Monitoring and control of transport networks using parsimonious models

Ruzanna Mat Jusoh

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Engineering
College of Science and Engineering
University of Glasgow

February 2019

My Lord! Truly, I am in need of whatever good that You bestow on me

28:24

Abstract

The growing number of vehicles on the roads coupled with inefficient road operations have generated traffic congestion. Consequently, traffic congestion increase trip time and indirectly contributes to poor quality of life and environmental pollution. Therefore, alleviating traffic congestion, especially in urban networks, is crucial and requires efficient traffic management and control. Recently, macroscopic operational scheme has become the preferred method for monitoring and mitigating traffic congestion due its simplicity in modeling complex large-scale cities and low computational effort. The schemes are based on parsimonious models known as Macroscopic or Network Fundamental Diagram (MFD or NFD) which provides an aggregated model of urban traffic dynamics, linking network circulating flow and average density. This thesis deals with an open problems associated with two main applications of NFD in transportation networks, namely: 1) Traffic monitoring and 2) Traffic flow control. Two parts of the thesis concentrates on each application separately.

The implementation of NFD in perimeter control strategy requires an accurate estimation of NFD where its measurements are reflected from sensors located at appropriate locations in the network. First part of the thesis elaborates a new approach for studying sensor selection for the development of operational or sparse-measurement NFD, with less number of sensor and associated measurements. An information-theoretic based framework is proposed for the optimal sensor selection across a transport network to assist an efficient model selection and construction of sparse-measurement NFD. For the optimal sensor selection, a generalised set covering integer programming (GIP) is developed. Under this framework, several tools to assess GIP solutions are uitilised. First, a correlation between variables is introduced as a "distance" metric rather than spatial distance to provide sufficient coverage and information accuracy. Second, the optimal cost of GIP problem is used to determine minimum number of sensors. Third, the relative entropy or Kullback-Leibler divergence is used to measure the dissimilarity between probability mass functions corresponding to different solutions of the GIP program. The proposed framework is evaluated with experimental loop-detector data of one week from central business district with fifty-eight sensors. Results reveal that the obtained sparse-measurement diagrams from the selected models adequately preserve the shape and the main features similar to a full-measurement diagram. Specifically, the coverage level of 24% of the network demonstrated the effectiveness of GIP framework. Simulation results also disclose the Kullback-Liebler diver-

gence as more generic and reliable metric of information loss. Such framework can be of great importance towards a cost-effective sensors installation and maintenance whilst improving the estimation of NFD for better monitoring and control strategy.

Second part of the thesis discusses the traffic flow control problem involving single input flow distribution from perimeter control strategy towards number of gated links at the periphery of the network. It if often assumed that input flow ordered by perimeter control strategy should be equally distributed to a number of candidate junctions. There has not been considerable research into limited storage capacity/different geometric characteristics at gated links as well as equity properties for driver. A control scheme for the multi-gated perimeter flow control (MGC) problem is developed. The scheme determines optimally distributed input flows for a number of gates located at the periphery of a protected network area. A parsimonious model is employed to describe the traffic dynamics of the protected network. To describe traffic dynamics outside of the protected area, the basic state space model is augmented with additional state variables for the queues at store-and-forward origin links of the periphery. The perimeter traffic flow control problem is formulated as a convex optimal control problem with constrained control and state variables. For the application of the proposed scheme in real time, the optimal control problem may be embedded in a rolling-horizon scheme using the current state of the whole system as the initial state as well as predicted demand flows at origin/entrance links. This part also offers flow allocation policies for single-region network without considering entrance link dynamics namely capacity-based flow allocation policy and optimisation-based flow allocation policy. Simulation results are carried for a protected network area of downtown San Francisco with fifteen gates of different geometric characteristics. Results demonstrate the efficiency and equity properties of the MGC approach to better manage excessive queues outside of the protected network area and optimally distribute the input flows. The MGC outperforms the other approaches in terms of serving more trips in protected network as well as shorter queues at gated links. Such framework is particularly of interest to city managers because the optimal flow distribution may influence the network throughput hence serves maximum number of network users.

Acknowledgements

In the name of Allah, the Most Gracious and the Most Merciful. Alhamdulillah, all praises to Him for the strength and His blessing upon completing this research study.

I would like to express the deepest gratitude to my supervisor, Dr. Konstantinos Ampountolas for his guidance, patience, motivation and extensive knowledge shared, over the past four years. I am honoured with his dedication in assisting me throughout the research and writing this thesis. Without his endless support and wisdom, it would be impossible to complete this research.

A very special gratitude also goes to my sponsors, Ministry of Higher Education Malaysia and National Defence University of Malaysia for providing me a golden opportunity to further pursue my doctorate education abroad not to forget, for the financial support throughout this postgraduate study.

I would also like to dedicate the sincerest recognition to the pillar of my strength, my family. My heartfelt thanks to my dear husband, Mohd Khuzairie bin Mohd Taufik for his undying prayers, sacrifice, moral and emotional support throughout this jouney. My parents, Mat Jusoh Husin and Asiah Taib and my parents-in-law, Mohd Taufik Radzali and Norrasidah Ariffin as well as my siblings (Nur Amalina, Mohd Haniff, Mohd Mustaqim, Nur Nasuuha and Mohd Fuad), they have provided me with a vital essence to my endeavour. I am truly grateful for their endless prayers, encouragement and support.

My acknowledgement would be incomplete without thanking my friends and other contributors who has been involved in completing this thesis. I would like to give a special token of appreciation to my colleagues in University of Glasgow and National Defence University of Malaysia for the help and constructive suggestions. All their support has turned this hardship into a wonderful and meaningful journey.

Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Ruzanna Mat Jusoh
School of Engineering
University of Glasgow

Contents

Abstract i

Acknowledgements iii

Declaration iv

Contents v

List of Figures viii

List of Tables xii

Nomenclature xiv

1 Introduction 1

1.1 Background and motivation . . . 1
1.2 Traffic monitoring . . . 2
1.3 Perimeter flow control . . . 3
1.4 Thesis outline . . . 4
1.5 List of publications . . . 5

2 Literature review 7

2.1 Introduction . . . 7
2.2 Macroscopic traffic flow modelling of cities . . . 7
2.2.1 Aggregated traffic at the city-level . . . 7
2.2.2 The MFD on urban networks . . . 8
2.2.3 Factors influencing the MFD . . . 11
2.2.3.1 Location of loop detectors . . . 11
2.2.3.2 Spatial distribution of link densities . . . 12
2.2.3.3 Traffic signal settings . . . 14
2.2.3.4 Traffic mode decomposition . . . 14
2.2.4 Modelling MFD with uncertainty . . . 14

2.3 Traffic monitoring with parsimonious model . . . 15
2.3.1 Introduction . . . 15
2.3.2 Sparse-measurement MFD . . . 17
2.3.3 Near-optimal sensor selection for traffic monitoring . . . 19
2.4 Traffic flow control of cities . . . 23
2.4.1 Introduction . . . 23
2.4.2 Perimeter traffic flow control . . . 24
2.4.2.1 Overview of existing control approaches . . . 26
2.4.3 Multi-gated perimeter flow control . . . 29
2.5 Concluding remarks . . . 33

3 Traffic monitoring of transport networks 35
3.1 Introduction . . . 35
3.2 Sparse-measurement NFD and selection of sensors . . . 36
3.2.1 Sparse-measurement NFD . . . 36
3.2.2 Optimal selection of sensors . . . 36
3.2.3 Kullback-Leibler divergence and model selection . . . 41
3.3 Results . . . 41
3.3.1 Data description and setup . . . 41
3.3.2 Full-measurement NFD . . . 42
3.3.3 Optimal sensor selection . . . 52
3.3.4 Ad-hoc strategies . . . 73
3.3.5 Comparative study . . . 73
3.4 Discussion . . . 79

4 Multi-gated perimeter flow control 80
4.1 Introduction . . . 80
4.2 Traffic dynamics of single-region networks . . . 81
4.2.1 Dynamics within the protected area . . . 81
4.2.2 Entrance link dynamics . . . 83
4.2.3 State-space model . . . 85
4.3 Multi-gated perimeter flow control via MPC . . . 86
4.3.1 Control objective . . . 86
4.3.2 Model-predictive control . . . 87
4.4 Practical perimeter-ordered flow allocation policies . . . 88
4.4.1 Introduction . . . 88
4.4.2 Capacity-based flow allocation policy (CAP) . . . 89
4.4.3 Optimisation-based flow allocation policy (OAP) . . . 90
4.5 Case study and results . . . 91

4.5.1 Case study description . . . 91
4.5.2 Perimeter flow controller design . . . 92
4.5.2.1 Geometric characteristics of gates . . . 92
4.5.2.2 Single-region control without entrance link dynamics . . . 93
4.5.2.3 Single-region control with entrance link dynamics: Multi-gated control (MGC) . . . 93
4.5.2.4 Performance assessment criteria . . . 95
4.5.3 Results . . . 97
4.5.3.1 Perimeter flow control without entrance link dynamics . . . 97
4.5.3.2 Multi-gated perimeter flow control: Open-loop control results 105
4.5.3.3 Multi-gated perimeter flow control: Closed-loop control results 113
4.5.3.4 Perimeter-ordered flow allocation policies: CAP & OAP . . . 133
4.5.4 Performance assessment . . . 147
4.5.4.1 Comparison between MGC and CAP/OAP . . . 147
4.5.4.2 Equity properties of MGC . . . 150
4.6 Concluding remarks . . . 151

5 Conclusions and future work 153
5.1 Overview . . . 153
5.2 Discussion and conclusions . . . 154
5.3 Future work . . . 157

A Information theory 158
A.1 Entropy . . . 158
A.2 Relative entropy and measure of information correlation . . . 159

B Model predictive control 162
B.1 Linearisation of nonlinear systems . . . 162
B.2 Model predictive control . . . 163
B.2.1 Basic components of MPC . . . 163
B.2.2 Unconstrained MPC . . . 165
B.2.3 Constrained MPC . . . 167

Bibliography 171

List of Figures

1.1 Thesis outline. . . . 6

2.1 Conservation law. . . . 8
2.2 Loop detector data in Yokohama ([34]). . . . 9
2.3 Time series of region average flow and outflow ([34]). . . . 10
2.4 Example of MFD in other cities. . . . 11
2.5 MFD in Toulose; inset: hysteresis for morning ([14]). . . . 13
2.6 Theoretical MFD ([3]). . . . 16
2.7 Applications of MFD. . . . 24
2.8 Single region. . . . 26
2.9 Multi region. . . . 26

3.1 The set covering problem. . . . 37
3.2 Schematic map of Chania, Greece [60]. . . . 43
3.3 Network-wide NFDs for Monday to Friday. . . . 45
3.4 Network-wide NFDs for Saturday and Sunday. . . . 46
3.5 Flow and occupancy time series of Monday. . . . 46
3.6 Flow and occupancy time series of Tuesday and Wednesday. . . . 47
3.7 Flow and occupancy time series of Thursday and Friday. . . . 48
3.8 Flow and occupancy time series of Saturday and Sunday. . . . 49
3.9 Network-wide full-measurement NFDs for Monday and Tuesday. . . . 50
3.10 Network-wide full-measurement NFDs for Wednesday to Sunday. . . . 51
3.11 Comparison between pmf and geometric distribution Monday and Tuesday. . . . 59
3.12 Comparison between pmf and geometric distribution Wednesday and Thursday. 60
3.13 Comparison between pmf and geometric distribution Friday and Saturday. . . . 61
3.14 Comparison between pmf and geometric distribution Sunday. . . . 62
3.15 Comparison between pmf and geometric distribution at $d^f \geq 1.0$ Monday. . . . 62
3.16 Comparison between pmf and geometric distribution $d^f \geq 1.0$ Tuesday and Wednesday. . . . 63
3.17 Comparison between pmf and geometric distribution $d^f \geq 1.0$ Thursday and Friday. . . . 64

3.18 Comparison between pmf and geometric distribution $d^f \geq 1.0$ Saturday and Sunday. . . . 65
3.19 Comparison Friday pmf. . . . 66
3.20 Kullback-Leibler divergence versus cost of the GIP Monday and Tuesday. . . . 66
3.21 Kullback-Leibler divergence versus cost of the GIP Wednesday to Sunday. . . . 67
3.22 (a) Information entropy (in bits) of each day of a week; (b) Kullback-Leibler divergence in function of δ for one week. . . . 68
3.23 Sparse-measurement NFD approximations Monday and Tuesday and different δ values. . . . 69
3.24 Sparse-measurement NFD approximations Wednesday and Thursday and different δ values. . . . 70
3.25 Sparse-measurement NFD approximations Friday and Saturday and different δ values. . . . 71
3.26 Sparse-measurement NFD approximations Sunday and different δ values. . . . 72
3.27 Partition of the network into five regions (A, B, C, D, and E). . . . 73
3.28 Fit line of different models. . . . 75
3.29 One thousand of replications involving twenty randomly selected links. . . . 76
3.30 NFD of less number of links for Friday according to different models. . . . 77
3.31 Selected link (sensor) locations according to different models. . . . 78

4.1 Protected network with entrance link dynamics. . . . 83
4.2 Protected network and controlled gates of entrance. . . . 91
4.3 Network fundamental diagram of Downtown San Francisco [1]. . . . 92
4.4 Version I: State and control trajectories for different initial states and weights. . . . 98
4.5 Version II: Control trajectories for gates 1–6. . . . 100
4.6 Version II: Control trajectories for gates 7–12. . . . 101
4.7 Version II: Control trajectories for gates 13–15; state trajectory of PN. . . . 102
4.8 Version II (constrained): Control trajectories for gates 1–6. . . . 103
4.9 Version II (constrained): Control trajectories for gates 7–12. . . . 104
4.10 Version II (constrained): Control trajectories for gates 13–15 and state trajectory of PN. . . . 105
4.11 (a) State trajectories of PN; (b)–(d): Relative vehicle queues at origin links 1–3 for different initial states with and without disturbances (open-loop control results). . . . 107
4.12 Relative vehicle queues at origin links 4–9 for different initial states with and without disturbances (open-loop control results). . . . 108
4.13 Relative vehicle queues at origin links 10–15 for different initial states with and without disturbances (open-loop control results). . . . 109

4.14 Control trajectories of MGC at gates 1–6 for different initial states with and without disturbances (open-loop control results). 110
4.15 Control trajectories of MGC at gates 7–12 for different initial states with and without disturbances (open-loop control results). 111
4.16 Control trajectories of MGC at gates 13–15 for different initial states with and without disturbances (open-loop control results). 112
4.17 Sensitivity analysis and performance of MGC for different optimisation horizon N_o and demand scenarios: (a) TTS within the protected network and outside network area; (b) RQB within the protected network and outside network area . 114
4.18 Medium demand senario; (a): Demand at PN; (b)–(f): Demand at origin links 1–7. 117
4.19 Medium demand scenario: Demand at origin links 8–15. 118
4.20 Medium demand scenario: Ordered flow for gates 1–6. 119
4.21 Medium demand scenario: Ordered flow for gates 7–12. 120
4.22 Medium demand scenario: Ordered flow for gates 13–15. 121
4.23 Medium demand scenario: Relative vehicle queues at origin links 1–6. 122
4.24 Medium demand scenario: Relative vehicle queues at origin links 7–12. 123
4.25 Medium demand scenario: (a)–(c) Relative vehicle queues at origin links 13–15; (d) State trajectory at PN. 124
4.26 High demand scenario; (a): Demand at PN; (b)–(h) demand at origin links 1–7. 125
4.27 High demand scenario: Demand at origin links 8–15. 126
4.28 High demand scenario: Relative vehicle queues at origin links 1–6. 127
4.29 High demand scenario: Relative vehicle queues at origin links 7–12. 128
4.30 High demand scenario; (a)–(c): Relative vehicle queues at origin links 13–15; (d) State trajectory of PN. 129
4.31 High demand scenario: Ordered flow at gates 1–6. 130
4.32 High demand scenario: Ordered flow at gates 7–12. 131
4.33 High demand scenario: Ordered flow at gates 13–15. 132
4.34 OAP: Ordered flow at gates 1–6. 135
4.35 OAP: Ordered flow at gates 7–12. 136
4.35 OAP: Ordered flow at gates 13–15. 137
4.36 OAP: Relative vehicle queues at links 1–6. 138
4.37 OAP: Relative vehicle queues at links 7–12. 139
4.38 OAP: (a)–(c) Relative vehicle queues at links 13–15; (d) State trajectory at PN. 140
4.39 CAP: Ordered flow at gates 1–6. 141
4.40 CAP: Ordered flow at gates 7–12. 142
4.41 CAP: Ordered flow at gates 13–15. 143
4.42 CAP: Relative vehicle queues at links 1–6. 144
4.43 CAP: Relative vehicle queues at links 7–12. 145

4.44 CAP: (a)–(c) Relative vehicle queues at links 13–15; (d) State trajectory of PN. 146
4.45 Total time spent within the protected network and gated links under MGC, CAP, and OAP for four initial states in a scenario without external demand. 148
4.46 Total time spent within the protected network area and gated links under MGC for different initial states and two scenarios with external demand. 152

A.1 Venn diagram showing the relationship between entropy (conditional, joint) and mutual information. 160

B.1 The rolling horizon idea behind MPC [12]. 165

List of Tables

1.1 Summary of contributions. 4

2.1 Summary of perimeter flow control in the literature. 30

3.1 Candidate covering models derived from the GIP. 40
3.2 Capacity flow and critical occupancy for one week. 42
3.3 Capacity flow and critical occupancy of full-measurement NFD for one week. . 50
3.4 Solution of the set covering problem, entropy, KL-divergence and parameter of geometric distribution for Monday and Tuesday and different δ^f values. 55
3.5 Solution of the set covering problem, entropy, KL-divergence and parameter of geometric distribution for Wednesday and Thursday and different δ^f values. . . 56
3.6 Solution of the set covering problem, entropy, KL-divergence and parameter of geometric distribution for Friday and Saturday and different δ^f values. 57
3.7 Solution of the set covering problem, entropy, KL-divergence and parameter of geometric distribution for Sunday and different δ^f values. 58
3.8 Comparison of different models involving less number of links for Friday. . . . 74

4.1 Different characteristics of entrance links and controlled gates. 94
4.2 Summary of different controller design for perimeter flow control. 96
4.3 Total time spent (TTS) within the protected network area. 115
4.4 Total time spent (TTS) and relative queue balance (RQB) at gated links. 115
4.5 Storage capacity-based ratios of the fifteen entrance links. 134
4.6 Average value of TTS at PN under No control, MGC, CAP, and OAP for different initial states in a scenario without external demand. 149
4.7 Average value of TTS at gated links under No control, MGC, CAP, and OAP for different initial states in a scenario without external demand. 149
4.8 Vehicle queues under No control, MGC, CAP, and OAP for different initial states in a scenario without external demand. 150
4.9 Average value of TTS at PN under MGC for different initial states and two scenarios with external demand. 152

4.10 Average value of TTS at gated links under MGC for different initial states and two scenarios with external demand. 152

Nomenclature

Abbreviations

CAP	Capacity-based flow Allocation Policy
CBD	Central Business District
FD	Fundamental Diagram
GIP	Generalised Integer Programming
IP	Integer Programming
KL	Kullback-Leibler
LP	Linear Programming
LQ	Linear Quadratic
MFD	Macroscopic Fundamental Diagram
MGC	Multi-gated Perimeter Flow Control
MPC	Model Predictive Control
NFD	Network Fundamental Diagram
NP	Non-deterministic Polynomial-time
OAP	Optimisation-based flow Allocation Policy
OD	Origin Destination
PFC	Perimeter Flow Control
pmf	probability mass function
PN	Protected Network
QP	Quadratic Programming

RMSE	Root Mean Square Error
RQB	Relative Queue Balance
TTS	Total Time Spent

Notations Chapter 3

$\mathcal{C}$	minimum-cardinality subfamily
δ^f	threshold of relevant flow correlation
δ^o	threshold of relevant occupancy correlation
d_{ij}^f	"distance" of flow correlation
d_{ij}^o	"distance" of occupancy correlation
j	sensor index
$\mathbf{P}(X = x)$	The probability that the random variable X takes the value x
p	parameter of geometric distribution
$\hat{p}$	parameter estimation of geometric distribution
r_{ij}	radius between sensor i and j (m)
R	threshold of spatial distance
ρ	correlation between two variables
χ_j	binary variable associated with sensor selection at link j

Notations Chapter 4

$\mathbf{A}$	state matrices
$\mathbf{B}$	control matrices
$\mathbf{d}$	vector of disturbance variable
k	discrete time index
ℓ_o	vehicle queue at controlled gate o
n	accumulation
N_o	prediction horizon

N_p	control horizon
o	gated link index
$\mathcal{O}$	Number of controlled gates
q_o	input flow at controlled gate o
$\mathbf{Q}$	positive semi-definite weighting matrices
$\mathbf{R}$	positive definite weighting matrices
T	sampling time
$\mathbf{u}$	vector of control variable
$\mathbf{x}$	vector of state variable

Chapter 1

Introduction

1.1 Background and motivation

Traffic congestion is a major problem in urban road networks, which is attributed to the increased number of vehicles on roads and inefficient road operations. It appears when demand exceeds capacity of a region or network and often characterised by high densities and long travel times. The resulting saturated traffic conditions lead to vehicle queueing and network flow reduction that negatively impacts quality of life, safety and environment.

Thus, there is a need to mitigate traffic congestion and manage its effects. This could be accomplished through increasing road capacity or optimising existing infrastructure. Due to limited space available in modern urban areas, infrastructure upgrading and extension would be costly. Therefore, it calls for designing appropriate traffic management policies that could utilise the existing road infrastructure which provide more practical solutions to improve the efficiency and mobility of road networks.

A well-planned city is assessed by its performance of serving good accessibility and mobility. The performance of road network is usually assessed by microscopic complex models at the link (road) and/or junction level. In an attempt to come up with measuring the performance of two-dimensional networks at a macroscopic level, a parsimonious model known as Macroscopic or Network Fundamental Diagram (MFD or NFD) is used. It primarily shows the relationship between average network flow (circulating or output) and density (or vehicle accumulation). The NFD provides an approach for researchers and traffic engineers to model, monitor and control traffic in an aggregated way without the need of detail link. Recently, it is becoming increasingly significant to further develop the NFD-based applications in order to alleviate congestion in urban network area. Accordingly, this thesis deals with the monitoring and control of transport networks to contribute to the research and development of advanced traffic management systems.

1.2 Traffic monitoring

The NFD facilitates traffic engineers in monitoring transport networks by providing aggregated network traffic dynamics (at the region or city level) so they keep informed of the traffic level and if necessary, action can be taken accordingly. The shape and properties of the NFD depends on the network topology and current traffic conditions in the network, as reflected by the measurements of a number of loop detectors (or other sensors) placed at appropriate network locations. Hence, sensors data reporting is crucial to the building and successful deployment of NFD-based traffic monitoring schemes.

To date, the topic of sensor selection (with corresponding measurements) in gaining information to construct the NFD is still under discussion. Particularly a problem to construct an *operational* or *sparse-measurement NFD* using less number of sensors. For example, due to financial constraints, decision to select the optimal sensors are critical. Related issue is what/where should be the sensor location that can be both cost effective and provide most useful information. Another issue is about evaluating the adequacy of the information provided by the selected sensors to reconstruct a well-defined NFD for control implementation. Existing research [56, 73] are limited to address minimum required network coverage without revealing the locations of the sensors while automated framework to select sensors is not present. The motivation of this part is in line with the discussion of [36] who highlighted the importance to allocate sensors to maximise monitoring at network level and [28] who pointed the importance of detectors location for probe penetration rates estimation in data fusion exercise. Therefore sensors should be selected with care via a systematic method to accommodate in the development of sparse-measurement NFDs, which involve less sensors (corresponding measurements) and are in principle less costly.

In view of the sensor selection problem and related issues, the first part of this thesis deals with network-wide monitoring using an NFD. This part specifically contributes on developing a sound methodology for the optimal selection of sensors across a transport network that includes an *information-theoretic based framework* for efficient model selection. To the best of our knowledge, the presented information-theoretic framework is proposed for the first time in the transportation community and relevant literature. For the optimal sensor selection, a *generalised set covering integer programming* (GIP) is developed. Under this framework, several tools to assess GIP solutions is uitilised. First, as oppose to conventional approach, a correlation between variables is introduced as a "distance" metric rather than spatial distance to provide sufficient coverage and information accuracy. Second, the optimal cost of the GIP problem is used to determine minimum number of sensors. Third, the *relative entropy* or *Kullback-Leibler divergence* is used to measure the *dissimilarity* between probability mass functions corresponding to different solutions of the GIP program. Lastly, root mean square error (RMSE) is used as an additional accuracy metric to measure the differences between fitted reference (nominal) NFD and the observed values of the GIP model. The proposed framework is evaluated with

experimental loop-detector of one week data that is gathered from central business district with fifty-eight sensors. The results demonstrate that the obtained sparse-measurement diagrams from the selected models adequately preserve the shape and the main features similar to a full-measurement diagram. Simulation results also disclose the Kullback-Liebler divergence as a reliable and effective accuracy metric. Such framework can be of great importance towards a cost-effective sensors installation and maintenance whilst improving the estimation of NFD for better monitoring and control strategy.

1.3 Perimeter flow control

Alleviating congestion through optimising the supply of existing infrastructure in a network has always been a practical traffic management. In particular, there has been an extensive literature and work done about and with a tool known as perimeter flow control to serve traffic in cities at a desired level and to increase the network flow performance (see e.g., [1, 23, 37, 46, 55]). The idea of perimeter flow control is to hold some traffic at the perimeter of the protected network (by longer red phases at traffic signal), so as to maximise the network throughput.

Regardless of the type of controller applied in perimeter flow control, the calculated optimal (single) input flow values to the protected network need to be assigned to a number of candidate signalised junctions located at the periphery of the network. Nevertheless, studies on perimeter flow control assume that a single input flow ordered by the perimeter control strategy should be equally distributed (or in a non-optimal way). Such a distribution policy applied independently to multiple gates of a protected network area would be efficient in a case of unconstrained origin link queues for vehicle storage. Gated queues at origin links must be restricted to avoid interference with adjacent street traffic outside of the protected network area. Moreover, for a suboptimal flow distribution, link storage capacity could be underutilised at some locations while some may face development of queues at origin links. Such condition brings imbalanced entrance for drivers from different gating links to enter the protected network. Thus, limited origin links storage capacity/ geometric characteristics and the requirement of equity for drivers are the main reasons towards the multi-gated perimeter flow control strategy.

With regards to suboptimal input flows, the second part of the thesis proposes a *multi-gated perimeter flow control* (MGC) scheme for congested transport networks using the parsimonious NFD. This work contributes on developing an integrated model for multi-gated perimeter flow control, which takes into account the traffic dynamics within the protected network area and outside of the protected area. The parsimonious NFD of urban networks is employed to describe the traffic dynamics of the protected network area. In describing traffic dynamics outside of the protected area, the basic state-space model is augmented with additional state variables for the queues at store-and-forward origin links at the periphery, similar to [21, 44]. The integrated model is then used to formulate a convenient convex or nonlinear optimal control problem

Table 1.1: Summary of contributions.

Part	Specific topics addressed	Main contributions
Traffic monitoring	Operational or sparse-measurement NFD	An information theoretic framework for near-optimal sensors selection
Traffic control	Multi-gated perimeter flow control	Optimal flow distribution at candidate gating links

with constrained control and state variables for multi-gated perimeter flow control. The aim is to equalise the relative queues at origin links and maintain the vehicle accumulation in the protected network at a desired point, while the system's throughput is maximised. The proposed framework is applied and evaluated to a protected network area of San Francisco with fifteen gates of different geometric characteristics. Results demonstrate the efficiency and equity properties of the proposed MGC scheme. The proposed approach manages the excessive queues outside of the protected network area and optimally distribute the input flows. It also fulfils the requirement of equity for drivers using different gates to enter a protected network area. Finally this part of the thesis offers two perimeter-ordered flow allocation policies, namely *capacity-based flow allocation policy* (CAP) and *optimisation-based flow allocation policy* (OAP), to facilitate the real-time deployment of perimeter-ordered flow strategies. Further, these flow allocation policies are benchmarked against the proposed multi-gated perimeter flow control.

The results of this research will help traffic engineers to develop advanced traffic management systems towards improving network throughput and directly serving maximum number of network users due to its optimal flow distribution across transport network. Concluding, this thesis offers solutions to two different problems, namely monitoring and control of transportation networks. Table 1.1 summarises the key contributions of this thesis organised in two main parts.

1.4 Thesis outline

The structure of the thesis is illustrated in Figure 1.1. The thesis has two main parts. Chapter 3 presents Part I (traffic monitoring) while Chapter 4 presents Part II (perimeter flow control). Chapter 2 is devoted to the literature review required for both aforementioned chapters. Precisely the rest of the thesis is organised as follows:

- Chapter 2 provides a comprehensive review on macroscopic modelling, monitoring and control of transport networks. First, the macroscopic modelling of cities (at network level) with parsimonious models is discussed and the macroscopic or network-wide fundamental diagram (MFD or NFD) of two-dimensional networks is revealed. Second, the applications of NFD on monitoring with particular focus on the sparse-measurement is discussed. Finally, the application of NFD to perimeter flow control is reviewed with particular at-

tention on the allocation of perimeter-ordered flows at multi-gated links.

- Chapter 3 develops deeper understanding of the first part of the thesis, where a generic integer-programming problem (set-covering problem) is developed for the selection of sensors. This problem is combined with an information-theoretic based framework for efficient model selection to reconstruct a sparse-measurement NFD, which preserves the main features of a full-measurement NFD.

- Chapter 4 develops an integrated model for the multi-gated perimeter flow control (MGC) problem. Model-based predictive control (MPC) in a rolling horizon framework is used to solve the corresponding constrained optimal control problem. A case study of the application of the proposed MGC and allocation policies CAP & OAP to the protected area of Downtown San Francisco, CA, is presented.

- Chapter 5 provides conclusions and suggestions for future research.

- Appendix A provides the necessary background on information theory to develop a measure of information correlation, which is used as a "distance" metric to provide sufficient coverage and information accuracy in a set covering problem in Chapter 3.

- Appendix B provides the necessary background on the linearisation of nonlinear systems and introduces model predictive control (MPC) for the constrained control of linear discrete-time systems. MPC is employed for the solution of the multi-gated perimeter flow control developed in Chapter 4.

1.5 List of publications

The following are publications arising from this thesis:

Conference papers

1. R. Mat Jusoh and K. Ampountolas, "Distributed perimeter flow control of transport networks," in *49th Annual UTSG Conference*, Dublin, Ireland, 2017.

2. R. Mat Jusoh and K. Ampountolas, "Multi-gated perimeter flow control of transport networks," in *25th Mediterranian Conference on Control and Automation (MED)*, Malta, 2017, pp. 731–736.

3. R. Mat Jusoh and K. Ampountolas, "Traffic monitoring on sparse-measurement network-wide fundamental diagrams," in *50th Annual UTSG Conference*, UCL, London, 2018.

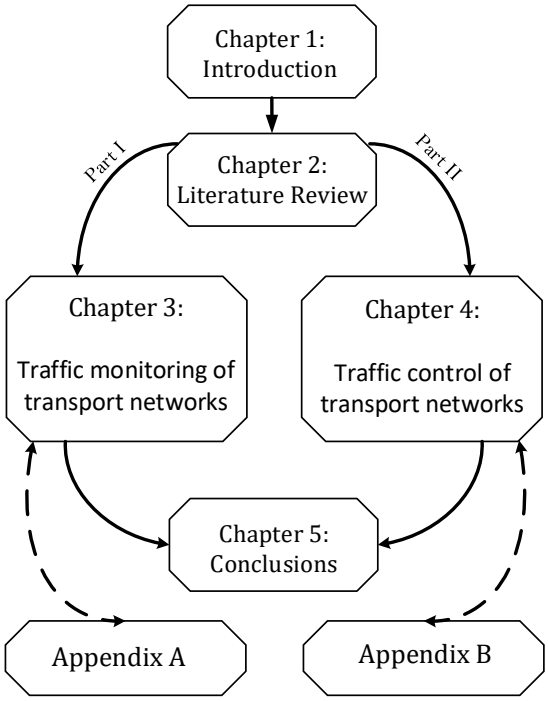


Figure 1.1: Thesis outline.

4. R. Mat Jusoh and K. Ampountolas, "Optimal selection of traffic sensors: An information-theoretic framework," in *2019 American Control Conference (ACC)*, Philadelphia, USA, 2019, *accepted/ to appear*.

Journal papers

1. R. Mat Jusoh and K. Ampountolas, "An information-theoretic framework for the near-optimal selection of sensors," To be submitted for possible publication to a journal, *In preparation*.

2. R. Mat Jusoh and K. Ampountolas, "Multi-gated perimeter flow control of monocentric cities," To be submitted for possible publication to a journal, *In preparation*.

Chapter 2

Literature review

2.1 Introduction

This chapter reviews relevant literature to develop greater insight into modelling, monitoring and control using aggregated model for urban networks. Section 2.2 provides the historical overview of macroscopic modelling of cities with parsimonious model, then it explores developments of the MFD. Section 2.3 presents an application of MFD in monitoring transport networks with a particular focus on sparse-measurement with less number of sensors. Besides, section 2.4 discusses the application of MFD in controlling the transport networks with specific attention on the perimeter traffic flow control at multi-gated links. Lastly, conclusions for this chapter is given in Section 2.5.

2.2 Macroscopic traffic flow modelling of cities

2.2.1 Aggregated traffic at the city-level

Various theories which described traffic dynamics at an aggregated level have been progressed over the past five decades. Known as macroscopic traffic models in urban networks, it describe the aggregate behaviour among traffic variables (mean speed, traffic flow and traffic density) at city (network) level. The traffic variables are represented in an aggregated way by summarising information of all vehicles, and it does not specify details of individual vehicles. Such concept of macroscopic model is of great importance allow for observing (monitoring) traffic dynamics at the network level, measuring the quality of traffic service in a city and further to perform necessary precautions such as alleviate congestion through traffic control.

Early works of an aggregated traffic were proposed in the past to model and understand macroscopic relationships at urban networks. An attempt has been made to describe the complexity of traffic flows in urban road networks began as early as in 1969 where [39] was the first to initiate the concept of macroscopic relationship between the mean speed in the centre part

Figure 2.1: Conservation law.

of a city and relate it with the accumulation of vehicles. This relationship shows an optimum accumulation but unable to verify the theory due to the lack of sufficient data at that time.

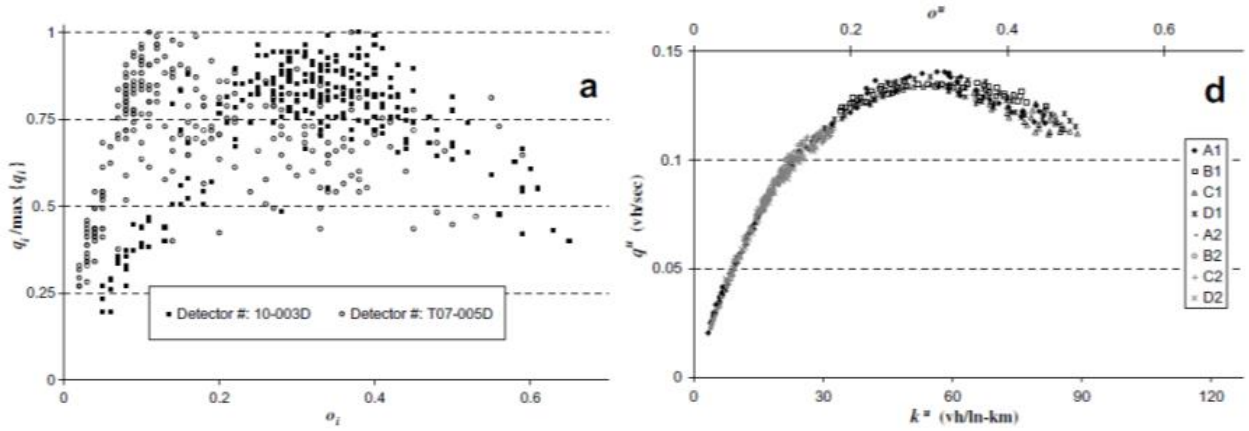
The macroscopic model of urban traffic further developed by [51] with the introduction to model traffic in a city by two-fluid model; one part representing the stopped vehicles due to e.g. incidents and congestions while the other part represents moving vehicles in the city. The model realistically relates the speed and density of an urban traffic network behind the physics philosophy where the average speed in urban network is a function of the fraction of stopped vehicles at any given time. [10] then undertook empirical validation of the two-fluid model using data experiment in Austin, Texas. Based on preliminary analysis, results postulated that average flow in a city could be expressed as product of speed and accumulation of vehicles. To enhance the understanding of the traffic in an urban area using the two-fluid model, [69] further investigated the network-level relationships by using simulation data. They observed the basic traffic flow variables: flow, density and speed with a fundamental relationship that flow is a product of density and speed. Results based on simulation confirm that it is possible to characterise traffic variable using macroscopic models and it is present at network level.

Although prior studies of macroscopic models showed remarkable developments regarding the urban aggregate traffic modelling, however, the studies looked for light traffic conditions or in data coming from simulations. It is not sufficient i) to describe congested conditions in an urban network change over time and ii) empirical evidence to demonstrate the existence of macroscopic models of urban networks is absent.

2.2.2 The MFD on urban networks

The whole idea of MFD-based modelling is about using aggregate network models, based on conservation law and outline relationship between space mean flow or outflow (veh/h) with vehicle density or accumulation (veh). The vehicle conservation equations as depicted in Figure 2.1. The principle is about the change in the number of vehicles (accumulation) in time equals the balance of input and output during the examined period.

Approach taken by [23] is among the pioneer who simplified the description of traffic dynamics. The author reintroduced a principle of MFD as an aggregate model of a city where it can be treated as few homogeneous reservoirs. The model required variables input and output which could be used without information of origin-destinations then a city can be expressed by a macroscopic function consists of variables flow, density and accumulation. The author claims



(a) flow vs. occupancy for two detectors

(b) average flow vs. average occupancy from all detectors

Figure 2.2: Loop detector data in Yokohama ([34]).

an MFD should exist for urban networks with uniformly loaded congestion if the traffic condition changes slowly with time. The study also observed that overcrowded networks in case of gridlock tend to have performance degradation, and suggested an adaptive control strategy to improve mobility and relieve congestion by controlling vehicle accumulation into the network.

Following the theoretical work by [23] led [34] to conduct field experiment using data from 500 loop detectors and successfully to be the first reported the existence of MFD in a large urban area in Yokohama, Japan. The constructed MFD confirms by aggregating the highly scattered plots of flow vs. density. MFD with very low scatter exists which links space-mean flow and density. They refer the aggregated variable weighted sum of flows of all links as "production" and weighted sum of link densities as "accumulation", both quantities are weighted with link lengths (lane-km). Figure 2.2 (a) illustrates data from individual detectors that is highly scattered and when aggregated the scatter nearly disappear and form along with a well-defined curve in Figure 2.2 (b). Instead of monotonic relationship between speed and flow which mostly found in early aggregated models, the MFD in Figure 2.2 (b) exhibit a concave like-shape with three special parts; free-flow, capacity and congested. The main feature is that for a (more or less) critical vehicle accumulation (or traffic density), flow capacity is reached; and, thus throughput is maximised. Figure 2.3 depicts an almost linear relationship between average network flow and its outflow (i.e., the trip completion rate, the rate vehicles reach their destinations). This relationship represents almost time-invariant value for the network average trip length. This criteria is important indicating average network flow can be observed with sensors (e.g., loop detectors) while outflow is more complicated and require full observation of network boundary and observation of ending trips [34].

The fundamental diagram can be reproduced either by using weighted or unweighted quantity. The weighted quantity take into account the link length while unweighted quantity do not

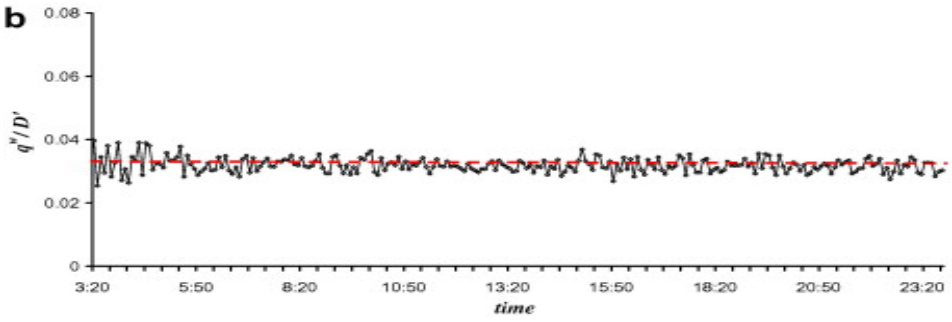


Figure 2.3: Time series of region average flow and outflow ([34]).

take link length. The formula of the quantity are given as [34]:

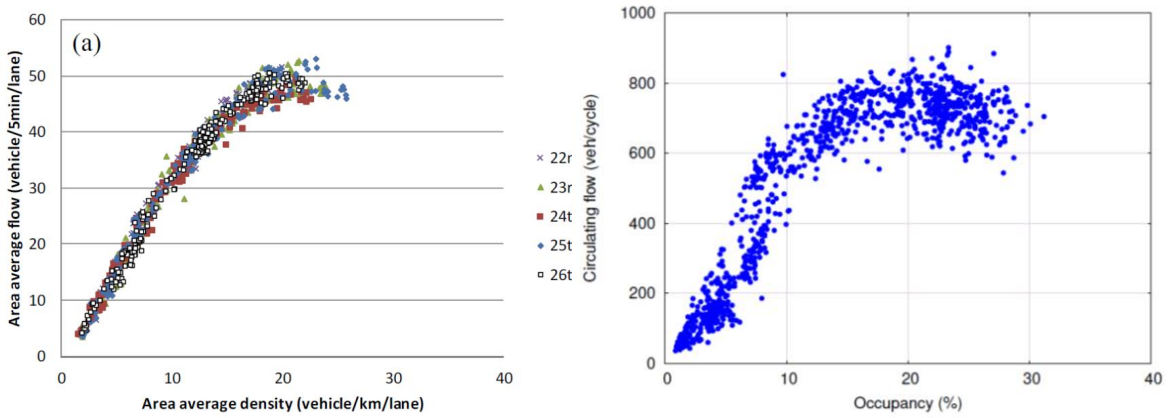
$$
\begin{aligned}
q^w &= \sum_i q_i l_i / \sum_i l_i, \\
q^u &= \sum_i q_i / \sum_i 1, \\
o^w &= k^w s = \sum_i o_i l_i / \sum_i l_i \\
o^u &= k^u s = \sum_i o_i / \sum_i 1
\end{aligned}
$$

where q^w is weighted flow, o^w is weighted occupancy, q^u is unweighted flow, o^u is unweighted occupancy, k is density, s is constant factor of effective vehicle length, q_i and o_i denote the flow and occupancy on link i with length l_i. It is also suggested by [34] the MFD can be used as a suitable perimeter control strategy to avoid oversaturated condition but serve traffic at a desired level. Since this prominent finding, the study of MFD received significant attention and plethora of works can be found.

For example, empirical study by [83] confirm the existence of MFD for signalised arterial networks in Brisbane, Australia. They used dataset from Bluetooth sensor. Figure 2.4 (a) illustrate the shape of MFD in Brisbane. Clearly, by aggregating the variables, the diagram shows a well defined relationship between flow and density with less scatter.

Another study by [7] verified that CBD in Chania, Greece exhibits an MFD. Using real data from 70 loop detectors, the urban area display a fundamental diagram with small level of scatters. The data were available at every 90 s for flow and occupancy and were aggregated over all-lanes data by calculating mean of lane occupancies and total summation of flows. Figure 2.4 (b) illustrates specific MFD on Monday where clearly show crisp relation between flow and occupancy. They also observed MFD of every day across one week has different shape and it is not unique.

A study by [24] provide theoretical ground and show the shape of MFD can be predicted using variational theory instead of using simulation. The curve derived from the theory fit well



(a) MFD in Brisbane ([83]). (b) MFD of Monday in Chania([7]).

Figure 2.4: Example of MFD in other cities.

with experimental data from Yokahama. The theory provides basic of analytical deviation of MFD by providing upper and lower bounds of the curve. Despite offering numerous advantages, the theory is limited to a network with regular topology (many route choice), slow varying demand and homogeneous network with similar links. Furthermore, the authors point out that obtaining a "well defined" (i.e. low scatter) MFD, depends on network topology and it is not universally expected.

Some example of empirical studies using real data which demonstrate the fundamental diagram exists in the real world has been discussed. The MFD exhibits a concave curve with a sweet point, where maximum average vehicle flow or throughput is observed. Although the MFD can well describe urban traffic dynamic, the existence of the fundamental diagram with less scatter is not universally expected. The key assumption of MFD is the homogeneity of congestion loaded across the network is not always expected in the real world. Several researchers have been challenged this assumption and investigated what can induce traffic heterogeneity.

2.2.3 Factors influencing the MFD

The curve of MFD is not unique and many factors such in the following can influence their shape and properties.

2.2.3.1 Location of loop detectors

A work by [14] was the first to argue homogeneity assumption for a well-defined MFD proposed by [34]. By using empirical data of city network in Toulouse, France, they showed examples of MFD when the network is heterogeneously distributed. Three causes of heterogeneity mentioned are i) type of network ii) presence of congestion and iii) location of loop detector (near or far from traffic light). The authors grouped the loop detector data according to their distance

to traffic light and then performed comparative analysis on a different network; i) highways vs. surface street networks and ii) center vs. suburbs. Based on the experiment performed, the distance between detectors and traffic signals has a strong impact on the slope of the MFD; the nearer the distance of detectors with traffic light, the lower the slope of MFD is. They further investigated the cause of scatter where they observed hysteresis in the MFD. High flow is observed in the onset and lower in the offset for the same density value as illustrated in Figure 2.5. Their conclusions are parallel with the claim made by [24] and added the distance between loop detector and traffic signal should be homogeneous to ensure a crisp curve of MFD.

Later, [19] set up few experiments of locating sensor location by using the different distance between loop detector and traffic signal. The reference pointed out the more quantity of the detector within the roads produce better MFD. They recommend that locating the loop detectors at different section of the roads (links) (e.g., upstream, downstream or middle) can cover different traffic conditions and hence create a better MFD. They also pointed out that without enough coverage from loop detector will exhibit uneven bias in MFD estimation.

2.2.3.2 Spatial distribution of link densities

In the interest of clarifying the key requirement for well-defined MFD, [71] was the first study to discover using a simulation that the spatial distribution of link densities gave a significant impact on the shape of MFD. They observed a network with heterogeneously distributed (variance of density over different location is high) would exhibit network flows smaller than those that approximately meet homogeneity conditions (low spatial variance of link density). They provide a reason that "an inhomogeneity in the spatial distribution of car density increases the probability of spillover, which substantially decreases the network flow".

A study by [36] further confirmed the effect of variability in space and time through an experiment on an arterial network in downtown Yokohama, Japan and freeway network in Twin Cities, Minnesota. The statistical experiments suggested the condition for a well-defined MFD is that "if the spatial distribution of link density is similar for two time interval with same number of vehicles in network, then the two time intervals have the same average flows". Thereby, the earlier assumption of homegeneity condition [34] is relaxed. While arterial network satisfied the condition, but it does not hold in freeway network where a non well-defined MFD is observed.

Spatial distribution of link densities is a crucial factor that affects the shape of a MFD. Evidently, for heterogeneous networks with multiple centres of congestion the observation of well-defined fundamental diagrams is more challenging. Networks with uneven congestion distribution may show traffic states that are well below the upper bound of a MFD and high scatter to line on a MFD. Based on the discussed findings, a possible solution to have a well-defined MFD is to clustering the heterogeneous network into a small number of homogeneous clusters with small variance of link densities.

An approach taken by [54] investigated the partitioning algorithm of the urban traffic net-

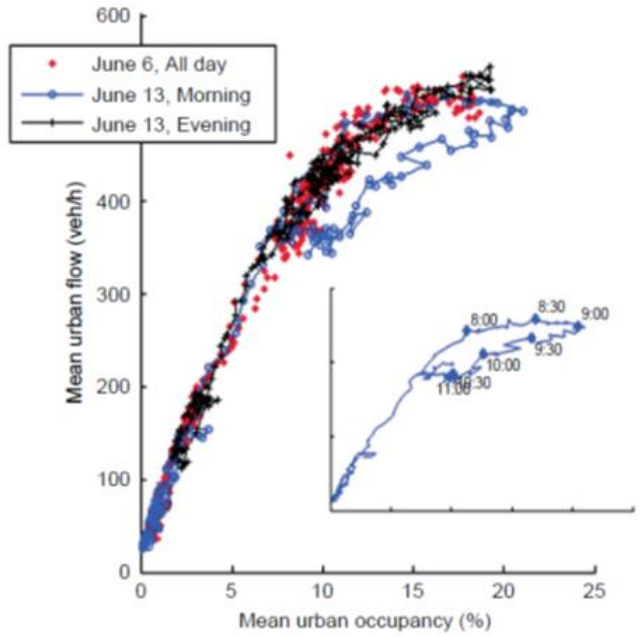


Figure 2.5: MFD in Toulose; inset: hysteresis for morning ([14]).

work in San Francisco. The work in [54] provide a foundation in clustering of heterogeneous transportation networks. The goal of partitioning is to obtain i) low variance of link densities within cluster and ii) spatial compactness of each cluster, which makes the potential for perimeter control application. Each goal is validated by density variance metrics and spatial compactness metrics respectively. The proposed algorithm can produce a partitioning with a desired number of clusters, demonstrated the effectiveness and is theoretically solid. However, the nature of congestion in transport network spreads and propagate dynamically over space and time features while the algorithm is limited to static traffic states. Hence, a dynamic partitioning is critical to ensure it works in urban networks.

The effect of partitioning is further investigated by [76] who proposed partitioning heterogeneous networks into connected homogeneous regions by obtaining directional congestion within regions. They suggested a new definition of link similarity based on running snakes and developed optimisation framework model and tested it on the real network in San Francisco and Shenzen. The 'snake' is referring to each sequence of roads and running snakes depicted by iteratively, one adjacent road is added based on its similarity to join existing added roads. The iteration continues growing by adding and absorbing different links near snake until all the links in the network are added. Later, [77] extends the investigation by capturing dynamic fashion and incorporate time factors over the traffic changes. The framework allows to identify links with the highest heterogeneity and re-cluster them with similar traffic patterns to minimise inhomogeneity effect.

In another way to alleviate the effect of inhomogeneity in networks, [58] derived a Generalised MFD which shows traffic production as a function of the average density in a network and

the spatial inhomogeneity of density across the entire network. They observed the Generalised MFD describes more of inhomogeneity in the production compared to MFD.

2.2.3.3 Traffic signal settings

One of the future work suggested in [14] is to investigate the effect of traffic signal upon the scatter of MFD. In general signal timing (cycle length, a proportion of green time) have an impact on MFD. The main idea is to have a proper timing of traffic lights across the perimeter of an urban area which will determine the inflow while keeping the traffic at maximal flow and prevent from saturated condition. An investigation by [3] explore this idea with their developed signal control strategies in city centre of Chania, Greece. Using scenario of strong demand with high fluctuations, they observed a hysteresis effect in the diagrams in the presence of different strategy. This means at the same accumulation value, network flow are different during onset (when the network is filled) and offset (when the network is emptied).

In [32], they studied the influence of adaptive signal control on network stability and MFD. The network is considered stable when they are in less congested state. The stability give an impact to homogeneous spatial distribution hence facilitate in producing relationships between traffic variables. Based on their observations, adaptive traffic signal contribute to stability only when a network is in medium congestion state. However, in a dense congested network, the adaptive traffic signal give irrelevant effect on the urban network and it should be supported with adaptive routing drivers. Work by [33] derived analytical theories for the shape of the MFD as a function of network and intersection parameters, using variational theory for different city topologies and signal structures.

2.2.3.4 Traffic mode decomposition

Unpredictability of traveller choices over route, time and mode of travel also affecting the shape of MFD. Particularly with different transportations mode (e.g., cars, buses) there is always competition for space in urban networks. A study by [38] taking into account the traffic mode decomposition in case of mixed multimodal traffic (cars and buses share same network infrastructure) and derived three-dimensional MFD model. They observed the network's circulating flow decreases with the number of buses serve in the network.

2.2.4 Modelling MFD with uncertainty

For a better representation of the traffic dynamics, there are few works which included the robustness concerning uncertainties into the MFD modelling.

Recent work [3] has identified that flow capacity in urban networks may be observed over a range of accumulation-values, and thus a critical accumulation cannot accurately be specified or is subject change due to adaptive signal control [3]. Therefore for a better representation

of the traffic dynamics, [7] included the uncertainty into the MFD-based model representation through tracking in real-time prevailing critical accumulation over a range of accumulations. The Kalman filter uses different increments while estimating the critical accumulation to allow the different rate of increase (or decrease) during the onset (or offset) of congestion to improve the tracking performance.

A study by [46] dealing with this problem by considering uncertainty in the value of trip completion flow that can differentiate between upper and lower envelope (interval) of MFD which capture the nonlinearity dynamics. Shortly after that, [42] continued the work and consider modelling of uncertain MFD with destination decomposition of the accumulations. Recently, [62] have introduced uncertainty traffic flow description with additional dynamic weights for a single region network. [8] has extended the network-wide uncertainty to multimodal networks for traffic mode decomposition.

As a result of this section 2.2, macroscopic modelling for the urban network from early works based on aggregated traffic is surveyed, how it developed to the known MFD, factors give impact on its shape to modelling uncertainty in MFD. Furthermore, macroscopic models offer number of benefits such: input needs for macroscopic simulation are reduced and large-scale city network modelling is easier, low computational effort, the low infrastructure required, produce more elegant control schemes and demand-insensitive relationship between accumulation and network flow [34]. In the following, the literature review is proceeded into two section which mainly discusses the application of MFD in monitoring and controlling the transportation networks. The first section describes the monitoring part in Section 2.3. This literature is related to Chapter 3. Following that literature in regards to control of traffic congestion in urban areas in Section 2.4 is described. The literature in this section is significant with Chapter 4.

2.3 Traffic monitoring with parsimonious model

2.3.1 Introduction

Recently, MFD has been found to be particularly useful for monitoring traffic congestion in urban areas [23, 34, 69]. MFD is used for monitoring by providing an update information of the network state status. The shape and regime (component) of the MFD can illustrate the traffic condition of a network. Figure 2.6 illustrates the theoretical shape of the MFD for urban road networks, where the vertical-axis represents total network flow (circulating flow or trip completion flow rate) while horizontal-axis represents number of vehicles in the network (proxy of occupancy or density). Occupancy is the percent of time a traffic loop detector is occupied by vehicles. Occupancy indicates spread of congestion across the traffic network.

It can be seen in Figure 2.6, as vehicle accumulation increased from zero, the network throughput increases to a maximum (flow capacity) and then turns down and decreases sharply

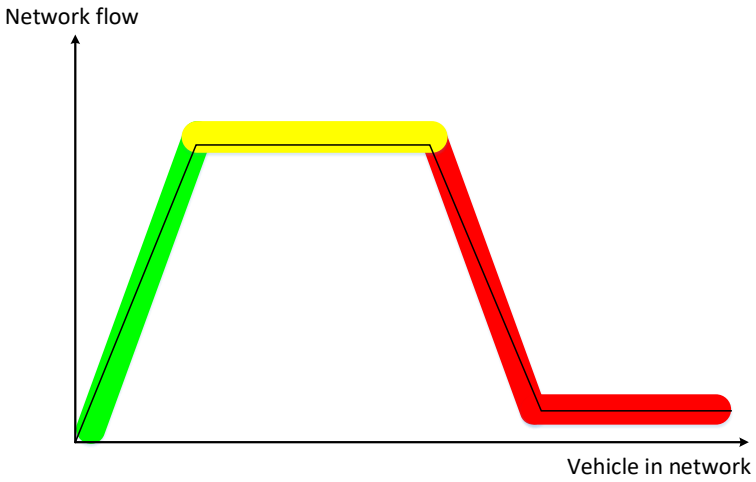


Figure 2.6: Theoretical MFD ([3]).

to a low value possibly zero or complete standstill (gridlock). The corresponding maximum point of flow capacity has the critical occupancy (or range of critical occupancy). The term critical occupancy is extensively used to distinguish the limit between normal and congested traffic situations. Figure 2.6 also clearly illustrate three important component in MFD. The description is as following; when the network occupancy is less than critical point, traffic state is regarded as uncongested conditions or free flow (green zone with incline gradient). The slope of the line is corresponding to the average speed in the network. When occupancy lies in the critical range the traffic state is regarded to be in the saturated conditions (yellow zone with horizontal line), while the traffic state regarded to be in the oversaturated phase (red zone with decline gradient) if the network occupancy exceeds critical point.

Through monitoring, MFD assists the traffic engineer by providing the network state (e.g., congestion level) information whether the network is performing at green, yellow or red zone of the MFD. Obviously, the red zone is an undesired traffic state. When an urban network is congested, queues take place which leads to flow deterioration, travel time increase and infrastructure is underutilised. Hence with this information, traffic controller will try to avoid the network to be in oversaturated condition but serve traffic at a desired level. Ideally, the network should operate at or below critical occupancy (accumulation) to have maximum flow condition. By proper traffic management, this aim can be achieved.

It is inevitable MFD is beneficial for researchers and traffic engineers but at the same time, the relevant data (or infrastructure/sensors) to construct an MFD are not always available. To account the data limitation, number of studies have examined different data sets. Typically, the information for MFD estimation comes from loop detector data obtained from fixed sensors or GPS-enabled devices probe vehicles and reports by cellular phones obtained from mobile

sources. In the following, a brief overview of studies concerning MFD estimation using fixed sensors or mobile sources is discussed.

A work by [65] presented a comparison study between speed from mobile data and flow estimates from detector data to observe the state of the urban network. They observed the estimated MFD does not require probe penetration rate (PPR) in advance. PPR is the number of equipped vehicles compared to the total number of vehicles in the area of interest [4].

Following that, [72] proposed a method to combine mobile probes data with loop detector data to estimate network-wide traffic variables. Mobile probes data serves a substantial source of traffic data which can provide information on vehicles' location, speed, and distance traveled. However, the assumption of probe vehicles data is homogeneously distributed across the network seem far from reality.

Taking into consideration regarding heterogeneously distributed of mobile probe data across the network, [28] estimated an MFD using mobile data. They proposed an algorithm to estimate the probe penetration rates based on k-clustering analysis. The obtained results highlighted on the importance of location and number of detectors on correct links as it is related with an optimum size of clusters to get better results.

An effort has been made to account both data sources (loop detector and mobile data) by [6] who propose a fusion algorithm to improve the accuracy of the estimated MFD. These works also showed that low penetration rates of probe vehicles reduce the accuracy of MFD estimation.

It has been shown that a low penetration rate of probe data in data fusion exercises can reduce the accuracy of MFD estimation. Another option is to combine data coming from fixed sensors and origin-destination or GPS trajectories in a resource allocation scheme to estimate the MFD [92]. This study revealed that the accuracy of MFD approximation is improved as fixed detection points are increased.

As shown in the literature reviewed above, whereas data from mobile sensors are widely-available throughout the network, the information obtained from loop detectors remains vital for the effective traffic monitoring of urban areas and the reconstruction of macroscopic traffic flow models. Our study limited to the data collected from fixed monitoring devices (loop detector data).

2.3.2 Sparse-measurement MFD

Recently, the estimation of the MFD with lower network coverage has gained interest. Simulation studies have confirmed that an operational or sparse-measurement MFD, which involves fewer sensors and corresponding measurements (reflecting different levels of network coverage), can be used for the efficient monitoring and perimeter control of congested urban areas.

In this direction, [56] used an MFD-based with less number of measurements, which they defined as reduced MFD and then the resulting MFD is exploited for solving gating control at city Chania, Greece. They compared the different level of network coverage, i.e. percentage from

total links in a network, for example, 5%, 10% and 35%. The corresponding links were chosen based on i) visual inspection during simulation specifically at a location where congestion starts to develop first and ii) personal judgment on what they think to be representative links. Results obtained using a fewer number of links were very encouraging; the controller works very well to protect the network and depicted equivalent performance as full measurement case (using all links). The accuracy of the obtained reduced MFD is evaluated in terms of total travel distance and total time spent.

Another study related to this topic is by [73] who studied the influence of location and number of measurement points for an MFD at the city of Zurich. They have developed a quasi-optimal strategy for link selection and compared its performance with some ad-hoc approaches, including random link selection or selection based on the distance to centres of congestion. The idea of ad-hoc strategies mainly follows traffic engineer experience for placing loop detectors in the network. The resulting MFD with less number of measurement is referred as partial MFD or pMFD. They observed the minimum level of coverage to be 25% of links used will guarantee a high accuracy. The accuracy is assessed by their proposed generalised method which measures the difference of density ratios between the two MFD (full and pMFD) while the metric is in percentage points (ppts).

Although these works are in the right research direction, however in both cases there is a lack of emphasis on providing a generic framework for efficiently building sparse-measurement operational MFDs. Unfortunately, from the existing method, the critical links are chosen depend on computer simulation, or personal judgment who must has a good background of the network, rely on quasi-optimal or ad-hoc blind strategies (involving random link selection), or subject to the standard practice of traffic community experience. The human judgment could be biased and this is a potential disadvantage. To make it more challenging, a large scale networks have multiple centers of congestion thus need a smarter way of selection with an automated and concise procedure, something not possible with a traditional, manual approach applied in current practice.

Although existing literature reveals about the minimum network coverage needed for a well-defined MFD estimation there is very limited study on incorporating the locations of corresponding measurements. If a city manager has budget limitations and wants to install some sensors on the roads, selecting an optimal location is very crucial at this stage. The strategic position of the traffic sensor is essential to obtain the most significant amount of information and ensure it is cost-effective. An optimal location with corresponding measurements will assist in estimating an accurate MFD which later facilitate in monitoring and control. Moreover, locating a sensor should not be location constrained where subjected to the specific area or level of congestion only. There is no guarantee flow at congested links might represent the real network flow. As mentioned by [19], the sensor (detector) should spread across the network to capture different traffic dynamics and provide overall traffic state conditions.

Furthermore, to evaluate the sufficient level of information from the sensors is another vital issue since monitoring resources are often scarce. The selected sensor should be the most informative about unsensed locations. Current accuracy assessment depends on i) visual inspection or ii) customised accuracy metric based on density ratios (in ppts). While it could be used to support evaluation, however, they are tailor-made to the problem and not generic. Hence, a reliable and systematic accuracy assessment of sparse-measurement would significantly benefit the development of traffic monitoring applications.

Therefore the optimal spatial resolution for loop detector placement and optimal temporal resolution for detector data reporting are crucial to the building and successful deployment of estimating the MFD and deserved an attention for further investigation. These studies motivated our work to develop a generic framework for efficiently building sparse-measurement operational MFDs, which explained in the following section.

2.3.3 Near-optimal sensor selection for traffic monitoring

Looking back at the research of the location problem, most of the studies model it into a conventional facility location problem which is used for determining the optimal locations of a hospital or gas station, etc., for serving the maximum number of customers while minimising the distance (e.g., average travel time or cost). The general problem is to locate new facilities or sensors to optimise some objective or cover a spatial area of interest or satisfy some demand points. Basic facility or location models include: set covering, maximal covering, p-center, p-median, fixed charge (where both the cost of building new facilities or locating sensors, and transportation costs are considered), hub, and maxisum [22, 25], [64].

Most popular models among facility location is the covering problem due to its applicability in real-world life, especially for service and emergency facilities. It seeks to ensure that each customer is “covered” (or served) by a certain facility if the distance between customer and facilities is equal or less than a predefined number (certain threshold, or required distance), sometime known as coverage distance or coverage radius [30].

Two classic type of the coverage models are set covering problem and maximal covering location problem. The set covering location problem concerns to locate minimum number of facilities while ensuring each demand can be served by at least one facility. It was first introduced by [50] who aimed to determine the minimum number of police needed to cover nodes on a network of highways. The mathematical model in set covering problems was later developed by [82]. They consider modeling the location of emergency service facilities such that every

demand points is covered by some facilities, as follows [30, 82]:

i : the index of demand nodes,

j : the index of facilities,

N_i : the set of potential locations within S so that $N_i = \{j | d_{ij} \leq S\}$,

x_j : a binary decision variable indicating whether the facility located at point j or not,

d_{ij} : the distance between demand node i and facility j, and

S : a maximum acceptable service distance. The model is as follows:

$$\min z = \sum_{j=1}^{n} x_j \tag{2.1}$$

subject to

$$\sum_{j \in N_i} x_j \geq 1 \qquad i = 1, \ldots, m \tag{2.2}$$

$$x_j \in \{0, 1\} \qquad j = 1, \ldots, n \tag{2.3}$$

Objective function (2.1) minimise the total number of located facilities. Constraint (2.2) shows the service requirement for demand i and constraint (2.3) states that all variables are binary. Meanwhile, the maximal covering location problem was proposed by [17] where it consists in locating p facilities that can cover the maximum amount of demand [79]. In addition, a review of covering-based problems in facility location can be found in [30].

In these mentioned models, a network is given with the locations of the demands to be served by the facilities and the locations of existing facilities or sensors. In general, these models work well in spatial facility location problems where the measure of distance and demand points are well defined. However, geometric assumptions related to a measure of distance are too strong in case of monitoring spatiotemporal phenomena, such as traffic flow in urban road networks by inductive loop detectors (traffic sensors). Traffic congestion usually propagates upstream in the network to random locations and traffic sensors make noisy measurements about the nearby regions, and this spatial sensing area is not usually characterised by a regular disk or a radius. For example, sensors located (far one each other) at arbitrary areas of the network can provide similar information and thus should be excluded. On the other hand, combining data from multiple sensors can give good predictions. In addition, the spatiotemporal distribution of congestion in traffic networks affects the shape of aggregated models used for traffic monitoring [36, 71]. Thus the notion of combination of data from multiple sensors in complex topological spaces is of fundamental importance and can not easily characterised by existing spatial models relying on the measure of distance. Except transport networks, the sensor selection problem arises in var-

ious other applications, including robotics [52], parametric identification of structural systems [74], wireless networks [89], and others [41].

In addition, the sensor location problem could be viewed from an information perspective. Sensors continuously provide information that helps characterise the status of the network. Obtaining data for the monitoring system is in first deals with maximising information. It means sensors are installed in network systems to acquire absolute accurate measurements (information). Information theory presented by [78] many years ago to determine maximum capacity of data transmission in communication systems. The past recent years have seen the development and application of information theoretic approach across different applications. Such work are in water level measurement [66], monitoring neutron flux distribution [81] and temperature measurements [61]. Information-based techniques are commonly adopted such as mutual information, information entropy and Fisher information. In [61], they observed that mutual information would work better than entropy alone as a criterion. They even observed the entropy criterion tends to place sensors along the borders of the area. As described above the sensor selection problem has been considered for various application except in transportation networks. Thus it is a motivation to apply the idea of information-theoretic framework together with location problem in monitoring the transportation networks. The information-theoretic framework able to provide a way to measure the amount of uncertainty of some quantities e.g. noisy measurements and highly unpredictable of traffic conditions while successfully quantify the quality and quantity of the collected information. Meanwhile the set covering problem is used since the method can provide the minimum number of sensors required together with the locations which ensure optimum level coverage. The combination method between information-theoretic framework and set covering location will provide a goodness in terms of where to strategically position traffic sensors in order to obtain largest amount of information on a network.

Research problem

The shape and properties of the MFD depend on the network topology and current traffic conditions in the network, as reflected by the observations of a number of loop detectors (or other sensors) placed at appropriate network locations. In our case, the full measurements is received from the sensing location placed at mid-block of links. From these full measurements, the MFD is created. In this work, one can see how the number of measurements (corresponding to number of sensors) is reduced and how one can reproduce the same MFD using limited number of measurements if it hold the same properties and preserves the original characteristics based on the MFD from full-measurements. In other words, this work are looking at the problem of selecting a subset of traffic sensors to be deployed in a network to characterise the spatiotemporal evolution of traffic density. This lead to an open question related to the optimal location of sensors selection:

1. How many measurements (all observation or subset of observations) in network are

needed to construct a well-defined MFD?

2. Where are the locations of these measurements to monitor?

3. How to select these locations to place the detector?

4. How to measure the accuracy of the resulting MFD?

Significance of study

This topic is worthy of study because the established mathematical optimisation will accommodate in choosing the best subset of possible sensors. Such approach can be helpful for improving and maximising the estimation of MFD and will facilitate traffic monitoring with sparse-measurement network-wide fundamental diagram.

In addition, the quality of traffic information heavily depends on the location and number of the sensors deployed on the road network. Poorly located sensor or unsuitable number of sensors can greatly increase installation and maintenance costs. It will also degrade the MFD estimation. Thus, how to use the minimum number of sensors to obtain high quality of traffic information is a problem worthy of study.

Selection of sensors are of fundamental importance not only due to physical or economic constraints but also to prevent redundancy of the same information which may occur between nearby sensors. Hence, it is very important to investigate and solve optimal spatial resolution for loop detector placement and optimal temporal resolution for detector data reporting to facilitate the construction of sparse-measurement MFD. It is expected that this topic will improve the limitation of existing methods and would hopefully trigger further developments in the macroscopic monitoring and control of urban road networks.

Research framework

In this work, the purpose is twofold. The first purpose is to develop and present a sound methodology for the near-optimal sensor selection across a network. The problem is modeled as a set covering integer programming problem, which is NP-hard (even with only polynomially many constraints, see e.g., [18]). A measure of correlation (based on Pearson correlation or mutual information) between random variables, reflecting a variable of interest, is introduced as a "distance" metric to provide sufficient coverage and information accuracy. As opposed to other sensor selection techniques, correlation is used between sensor locations rather than spatial distance to measure dissimilarity. The ultimate goal is to select sensors that are most informative about unsensed locations. The problem of finding the configuration that maximises mutual information is NP-complete.

Secondly, this study present an information-theoretic based framework for efficient model selection. The Kullback-Leibler divergence is used to quantify the approximation error incurred

between probability mass functions corresponding to different solutions of the integer programming. The proposed framework is applied to the problem of constructing sparse-measurement network-wide fundamental diagram using empirical inductive loop-detector data of one week from a central business district with fifty-eight sensors.

Contributions

The contributions of this topic can be summarised as follows:

1. An automated procedure for the sensor selection problem specifically in transport network is offered in order to construct an operational or sparse-measurement network-wide fundamental diagram.

2. An efficient strategy is formulated for efficient model selection which is based on information-theoretic approach.

Details of this topic will be discussed further in the next chapter. Chapter 3 develops a set covering integer programming formulation for the optimal selection of sensors and discusses how the proposed information-theoretic framework can be adopted, illustrates the effectiveness of the proposed framework with the use of real empirical data. Meanwhile, Appendix A presents some background in information theory, including entropy, Kullback-Leibler divergence and mutual information (see [78] for a detailed treatment). In the next section, existing literature for second part of the thesis will be discussed.

2.4 Traffic flow control of cities

2.4.1 Introduction

Apart from monitoring, MFD also has been extensively studied for the past few years to design traffic control schemes, to improve urban network performance and alleviate traffic congestion. An example of such a scheme is the perimeter traffic flow control.

MFD assist the traffic engineer by providing the network state (e.g., congestion level) information for the traffic controller. The idea of using the macroscopic relationship between network traffic variables is for optimising the accumulation of vehicles in an urban region. Returning to Figure 2.6, when the network or protected area is performing at low congested level (yellow zone), the city still has the ability to allocate incoming vehicles. At the green zone, the protected area is reaching its capacity and some actions are desirable to limit the access to the area to ease the growing number of vehicles. In the red zone, the network is congested or oversaturated congestion. This condition is beyond the network capacity and vehicles are no longer permitted to enter the protected area.

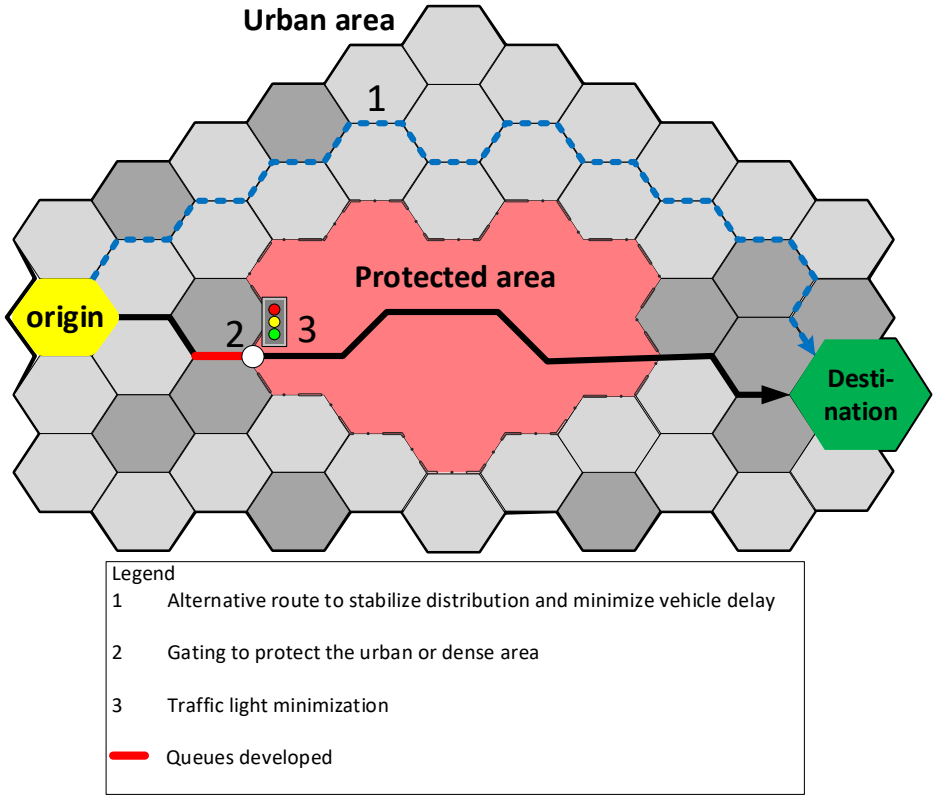


Figure 2.7: Applications of MFD.

In principle, the traffic controller aim to serve traffic at desired level. The network should operate at or below critical occupancy (accumulation) to have maximum flow condition. This can be achieved particularly through proper traffic management such as optimising traffic signal [3, 26, 88], protect the dense network through gating [1, 37, 55] or balancing the network traffic flow distribution through route guidance [48, 59, 87], congestion pricing [35, 91] and parking pricing [90]. Potential application of MFD in urban network is summarised in Figure 2.7.

2.4.2 Perimeter traffic flow control

The general idea of a perimeter flow control (or gating) strategy is to "meter" the input flow to the system by using appropriate traffic management such as traffic signals to hold vehicles outside the controlled area if necessary, so the congestion inside is prevented hence maximise the throughput.

An example of real world use of gating is Nottingham Zones and Collar experiment in 1975-1976 [84]. This experiment applied gating scheme which limiting access to the city centre using traffic signals timings. The "zones" are referring to residential areas from which traffic access to the main roads leading into the city while the "collar" is referring to traffic appeared around the central area due to delayed at traffic signals. Traffic is only allowed to enter the city centre area at the same time public transport is given priority. The objectives are to motivate non-essential

traffic to avoid the inner city area, and promote public transport by improving bus travel times relative to travel times by car. The report outlines on various aspects of implementation such as zone exit controls, collar controls, bus priorities and park and ride as well as how drivers responded to such scheme. It was reported that cars experienced excessive queues at particular gating points which were not designed to cope with queues.

Widely used strategies of responsive systems for urban traffic control such as Split, Cycle and Offset Optimization Technique (SCOOT) system [53], [13], [86] and Sydney Coordinated Adaptive Traffic System (SCATS) [67] are applicable to large-scale networks and employed to combat congestion of sensitive links or areas susceptible to congestion. SCOOT has been applied to more than 150 cities in UK and elsewhere [27]. This method is supplied with real measurements of traffic volume and occupancy (proxy to density) and is run repeatedly in real time to analyse effect of changes of splits, offsets and cycle time at individual intersections. Only the beneficial changes are submitted to controllers [27]. Although SCOOTS and SCATS are appropriate to help prevent under saturation congestion, their performance are deemed less efficient under severe congestion especially during peak period.

There have also been trials of gating in a project known as PRIMAVERA where it implemented two urban traffic control strategies namely UK SCOOT and Italian SPOT [31]. Both real time systems were improved so they can implement new algorithms and models in order to represent the reality more accurately. For example, SPOT systems has been modified to auto-gating which is able to store vehicle upstream of a critical bottleneck by directing green times according to the downstream queues or space left. The project focuses on integrated traffic control on urban arterial corridors. It has developed methods for queues management, public transport priority, protecting the environment and improving safety through traffic calming measures. Field trial in Leeds shown that with PRIMAVERA adoption, about 10% reduction in travel times for buses is achieved when contrasted with existing urban traffic control system.

More recently, an alternative approaches in regards with the idea of gating is reinvented using other tools for example the MFD. In an overview, most of existing literature in regards to MFD-based modelling using perimeter flow control has been fall into two main interest topics; single- region and multi- region.

In the case of a single region city with homogeneous traffic, the MFD exhibits a (more or less) unimodal and low-scattered shape and thus it indicates a critical accumulation value where maximum throughput is observed. MFD-based perimeter flow control has been proposed for single region cities in [23], [55], [45], [46], and [43]. Since the maximum throughput in the MFD is observed for a (more or less) critical value of vehicle accumulation, most of the recently developed perimeter flow control schemes for single region cities rely on a pre-specified set (or critical) value. Figure 2.8 show scheme involved for single-region network.

As the network-wide control has gain relevance, more complex concepts have been developed to extend the perimeter flow control for multi-region cities. Examples for multi-region can

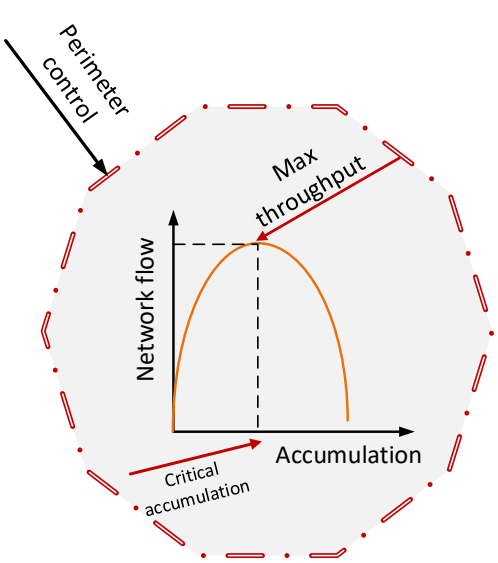


Figure 2.8: Single region.

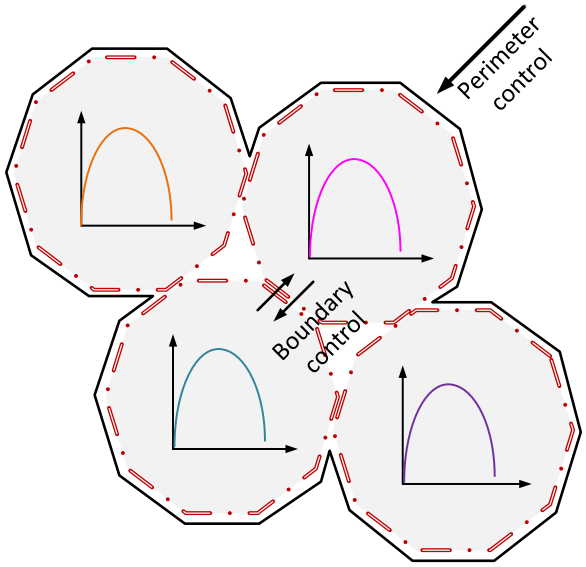


Figure 2.9: Multi region.

be found in [45], [37], [1], [49] and [75]. In the multi-region case (a heterogeneous city with many attractions of congestion) might not have a well-defined MFD. Therefore, it is important to partition or cluster the network into a number of homogeneous regions (see e.g. [54]). In this case, it is assumed that each homogeneous region exhibits a well-defined MFD.

In general, there are two types of scheme involved as illustrated in Figure 2.9. A perimeter flow control scheme is responsible for controlling the perimeter of the network, while a boundary flow control scheme located at the border between regions is responsible for manipulating the inter-flow transfers between neighbour regions. Perimeter and boundary flow control strategies aim at distributing the accumulation in each region as homogeneously as possible, and maintaining the rate of vehicles that are allowed to enter each region around a desired point while the system's throughput is maximised and minimise the total travel costs of all regions [1]. Researches made by [1], [37], [38] and [75] have studied the effect of heterogeneity in multi-region cities and designed perimeter control strategies that can decrease congestion heterogeneity, hence increase network performance.

2.4.2.1 Overview of existing control approaches

It has been shown in the literature that MFD-based control strategies, the perimeter flow control are effective in regulating global traffic performance. Researchers in urban traffic management have pursued various approaches. Some of the existing control schemes are reviewed briefly and among them are bang-bang control, feedback control approaches, Model Predictive Control (MPC) or rolling horizon schemes, robust control approaches and switching control schemes.

Bang-bang control

Studies by [23] employed a bang-bang control scheme to manage the accumulation to maintain the flow in the city on its maximum. The policy control is restricted to be between lower and upper bound. For example, if the condition in the network is stable and uncongested, the vehicles are allowed to enter the protected area up to the maximum allowable inflow. However, when the network operates in congested condition particularly when accumulation reaches its critical value, the inflow is turned to the minimum boundary. Nevertheless, bang-bang control is challenging when the maximum and minimum inflow is too extremes or too much oscillation hence it might not operate efficiently.

Feedback control

Proportional (P) and Proportional-Integral (PI) controllers, which are derived through the Linear Quadratic (LQ) optimal control theory, are employed in [55] and [1]. [55] used MFD-based modeling via application of gating control to improve mobility in saturated traffic conditions. In [1], a multivariable linear quadratic state feedback controller is studied for perimeter control. In both works, the nonlinear traffic dynamics are linearised to a pre-specified set point. The designed controllers do not include prediction of future or demand. They also imply control constraints indirectly after the design process.

Model predictive control

MPC is a control method that uses a model to predict future demand over an optimisation horizon. It is a receding horizon scheme, where at each time step an optimal open-loop of the problem with finite horizon is optimised, then only the first controller is applied to the plant and the procedure is carried out again. A receding horizon framework capable of handling state and control constraints which is an advantage in control design. For more detailed information on MPC, interested reader is referred to [40, 68].

A study by [3] implemented a rolling-horizon (model-predictive) control scheme to overcome the problem of signal control at the city centre of Chania, Greece. The control aim is to minimise oversaturation and balance the link queues at the road network to minimise the risk of queue spillback. The state variable involved is number of vehicles within link (queues) while the control variable is green time at a junction which reproducible based on a relationship between input flow, green time, saturation flows and cycle time. However, the proposed method remove queues after oversaturation happened and this limit the control impact.

Applying MPC in MFD-based modelling, [37] used perimeter control approach for two regions of a well-defined MFD. The optimal perimeter control is solved by model predictive control. The control aim is to maximise the output (trip completion flow of vehicles able to reach their destination) of the networks. Four state variables are corresponding to the inner and outer

region while two control variable are placed at the border of the regions to regulate the transfer flow within. The state and control constraints are introduced to respect lower and upper bound.

In an attempt to combine mixed network, [47] extended the model for the combination between urban and freeway network which means incorporating the ramp metering control strategy with the perimeter control. A centralised controller is achieved by the MPC scheme, while a cooperative decentralised control structure is developed to handle the coordination between urban and freeway management. The perimeter control on the boundary of regions manipulates the inter-transfer flows among regions, while on-ramp controllers considered the traffic flow distribution from urban areas to the freeway.

In a similar direction, [75] introduced region- and subregion- MFD based to explicitly model the heterogeneous density effect. They proposed a hierarchical perimeter flow control framework based on MFD-based modelling which implemented at the boundary of multiple sub-regions. At upper-level, MPC is used to identify the inter-transfer flows between multiple heterogeneous regions in the network. At the lower level, feedback control is used aiming at homogeneity spatial distribution congestion.

A work by [21] used nonlinear MPC in order to handle nonlinear nature of the controlled system where nonlinear function of the complete operational fundamental diagram is presented. Furthermore in the cost function contains performance criteria and input oscillation weighting which are different from other works. They compare their developed model with feedback control and the results outperform the feedback control.

Our work focuses more on the application of existing MPC design and the model involved is presented in Chapter 4. The detailed treatment of MPC is explained in B.2.

Robust control

As the network-wide control has increasingly gained relevance, more complex concepts have been developed hence recent works have focused on the robust control for urban traffic networks.

Robust controller which encompass worst-case scenario deals with uncertainty is designed to function at all network state without oblige to operate at a specific reference point. More precisely, [46] have considered accumulation uncertainty with an upper and lower envelope of the MFD, which can capture conservatively nonlinear dynamic phenomena. They designed a robust PI controller that stabilises the network system against all uncertainty for a single region network. Results showed that with a Linear Parameter-Varying (LPV) model, it could replace the original nonlinear model and a robust control scheme can be developed on the principles of Quantitative Feedback Theory.

In another work, [42] examine control of uncertain MFD by designing a robust interpolation-based constrained control scheme for multi-region networks. This scheme takes into account control and state constraints but does not integrate any optimisation criteria. In [62], a robust control scheme is developed to capture disturbances and to track in real-time the critical vehicle

accumulation. A constrained optimisation algorithm is used to distribute and allocate the input flow in an optimal way to a number of candidate junctions along the perimeter of the single region. Recently, [9] considered Proportional-Integral (PI) robust perimeter flow controller in dealing with the uncertainty of mix mode traffic.

Switching control

In an extra level of control to account more complex dynamics, [49] extended work in [37] towards twinning the perimeter and switching plans control to an urban network composed of several regions to solve the optimal traffic control problems. The perimeter control located at the border region control the transfer flow between region while the switching controller influences the dynamics by switch the timing plans of urban regions.

For a short survey, Table 2.1 summarise some relevant literature on MFD-based perimeter flow control; data or network involved, either in single- or multi-region or either the model developed taking into account the parameter uncertainty and queues at the exterior of the network.

2.4.3 Multi-gated perimeter flow control

Regardless of the control strategies applied, the perimeter control input (i.e. optimal flow) allowed to enter the network has to be distributed to the local intersections (or gated links). Recent studies on the perimeter flow control [1, 37, 42, 46, 49, 55] assume that a single ordered input flow obtained from a perimeter control strategy are equally distributed to be imposed at number of candidate junctions (or gated links) at the periphery of the network. Such control schemes do not allow for the direct consideration of the state and control constraints of the corresponding problem and usually employ suboptimal methodologies for the distribution of the obtained flows to the periphery.

Moreover, since the aim of perimeter control is to protect the inner region of the network, the access restriction of input flow at boundary of the network might create developing queues outside the region. Together with suboptimal flow distribution will lead to over-saturated back spilling and reduce the network performance.

In light of this, few works get an exception regarding assigning flows to different junctions (or gated links). For example in [62], they used optimal quadratic allocation algorithm to distribute the ordered flow from their Linear Parameter Varying (LPV) controller to number of candidate entrance gates. The ordered flow is equivalent to optimal link green time which is subjected to minimum and maximum bounds of cycle times. The objective function in their allocation algorithm is to minimise the difference between the ordered flow and (real) gated flow implemented at junctions. Although they try to take into account the green time (control) constraint, the state constraint is not in consideration.

A study by [21] have attempted to handle the gating problem of urban network in downtown

Table 2.1: Summary of perimeter flow control in the literature.

Study	Data (Network)	Objective	MFD Modeling			MFD Control	Queue outside region
			Single Region	Multi Region	Uncertainty		
[3]	Simulation (Chania)	Minimize link queues	✓			MPC	
[55]	Simulation (Chania)	Set point control	✓			PI[1]	
[46]	Simulation	Set point control	✓		✓	RPI[2] & SPI[3]	
[21]	Simulation (Budapest)	Maximize throughput	✓			NMPC[4]	✓
[62]	Simulation (San Francisco)	Set point control	✓		✓	LPV [5]	
[7]	Real (Chania)	Set point control	✓		✓	Adaptive	
[57]	Simulation (Chania)	Set point control	✓			PI	✓
[43]	Simulation	Maximize throughput	✓			Optimal control	
[1]	Simulation (San Francisco)	Set point control		✓		LQR[6] & LQI[7]	
[37]	Simulation	Maximize throughput		✓		MPC [8]	
[47]	Simulation	Minimize delay		✓		MPC	✓
[42]	Simulation	Set point control		✓	✓	LMI[9] & LP[10]	
[75]	Simulation	Minimize delay		✓		Hierarchy control: MPC & FHC[11]	
[49]	Simulation	Minimize time spent		✓		Hybrid control: MPC & SST[12]	
[44]	Simulation	Maximize throughput		✓		Optimal control	✓
[9]	Simulation (San Francisco)	Set point control		✓	✓	RPI[13]	

[1] Proportional Integral
[2] Robust Proportional Integral
[3] Standard Proportional Integral
[4] Nonlinear Model Predictive Control
[5] Linear Parameter Varying
[6] Linear Quadratic Regulator
[7] Linear Quadratic Integral
[8] Model Predictive Control
[9] Linear Matrix Inequalities
[10] Linear Programming
[11] Feedback Homogeneity Controller
[12] Switching Signal Timing
[13] Robust proportional Integral

Budapest. They enhanced the system model from [55] and allocated additional state variables to describe the queue dynamics at the protected network gates and solve the perimeter control problem using nonlinear model predictive control (NMPC). However, nonlinear MPC suffers from the computational effort needed to solve in real-time optimisation problem.

Another approach of this concept also worked by [44] who studied the effect of intensive vehicle assemble at the boundary of the network. A boundary optimisation control method for two-region networks is proposed to avoid queue interference and maintain the existence of MFD.

A preliminary results of the analysis has been shown in [57]. They proposed a queue management strategy to balance of relative queues formed at the gated links at perimeter of protected area. A relative queue is measured by dividing current queue length by its link's maximum queue length (maximum storage). The problem is formulated using quadratic knapsack optimisation problem.

Although some initial efforts have been made to consider the queue length at the perimeter, the distribution of allowable perimeter control input (i.e. optimal flow) to a number of candidate junctions at the periphery of the network has not been fully considered. Existing control schemes do not allow for the direct consideration of the different geometric characteristics of origin links, i.e., length, storage capacity, cycle time and other state and control constraints of the corresponding problem and usually employ suboptimal methodologies for the distribution of the obtained flows to the periphery, which is a potential disadvantage. Such a distribution applied independently to multiple gates of a protected network area would be efficient in case of unconstrained origin link queues for vehicle storage. However, gated queues must be restricted to avoid interference with adjacent street traffic outside of the protected network area and geometric characteristics of the different gates must be taken into account in the optimisation. The equity properties of perimeter flow control have not attracted considerable attention in the literature, although it is an important characteristic of any practical perimeter flow control application. These shortcomings call for new developments aiming at extending perimeter flow control strategies. Thus, limited origin links storage capacity/geometric characteristics and the requirement of equity for drivers using different gates to enter a protected area are the main reasons towards multi-gated perimeter flow control.

Overview of research

In this framework, this PhD research studies problem of designing efficient control schemes for the optimal distribution of input flows across the perimeter of single regions or the boundaries of neighbour regions. A multi-gated optimisation-based schemes is proposed that allows for the consideration of different types of state and control constraints.

Significance of study

This study will benefit from the development of more realistic schemes for the distribution of the input flows to the corresponding candidate junctions. The obtained distributed input flows can be appropriately transformed to prices since perimeter control can be viewed as a traditional zone control scheme with pricing (see e.g. the zone control scheme in London). Thus, effective pricing models and policies can be directly incorporated into the perimeter flow control problem to avoid queues and delays at the perimeter of a controlled area. The similar policies can be also utilised for toll station pricing. This is of great importance to practitioners and city managers for optimising the network capacity and serving the maximum number of travelers.

Moreover, a control strategy should be able to respect all constraints, minimise an objective criterion such as total time spent in network and consider an equity aspect for users of the different part in the network. Such an approach of appropriately distribute the input flow to the gated links responding to the needs of individual drivers using different gates to enter a protected area as they benefit from the equity policy applied in the scheme. It will also serve to balance the queues at links in an effort to reduce the risk of queue spillback and prevent infrastructure underutilisation.

Research framework

In this work, an integrated model for the multi-gated perimeter traffic flow control problem in urban road networks is proposed. The renowned network or macroscopic fundamental diagram of urban networks is employed to describe the traffic dynamics of the protected network area. To describe traffic dynamics outside of the protected area, the basic state-space model is augmented with additional state variables for the queues at store-and-forward origin links at the periphery as also in [21], [44]. The integrated model is then used to formulate a convenient convex or nonlinear optimal control problem with constrained control and state variables for multi-gated perimeter flow control. The aim is to equalise the relative queues at origin links and maintain the vehicle accumulation in the protected network around a desired point, while the system's throughput is maximised. This scheme determines optimally distributed input flow values (or feasible entrance link green times) to avoid queues and delays at the perimeter of a protected area while system's output is maximised.

Contributions

The contributions of this topic can be summarised as follows:

1. Developed a scheme which determines the optimal distribution of input flow for a number of gates located at the periphery of a protected network area with respect to link's storage capacity/ geometric characteristics.

2. Provide a scheme that ensure equity for drivers to enter the protected network area from different gates.

2.5 Concluding remarks

An overview of three important components; modelling, monitoring and control of transportation networks using MFD is discussed.

In Section 2.2 the history of macroscopic modelling of urban networks is surveyed from early works based on aggregated information at city-level. The reported earlier models emphasis more on a monotonic decreasing relationship between traffic variables and it is limited under less congested regime. The models then develop into two-fluid model, and to recent findings based on input-output system description known as MFD. These diagrams primarily show the relationship between average network flow and vehicle accumulation or traffic density and able to explain under uncongested and congested condition. They exhibit a concave like-shape and their main feature is that for a (more or less) critical vehicle accumulation flow capacity is reached; and, thus throughput is maximised. Successful to show MFD exist using real empirical data has encouraged researchers to develop MFD-based models, which is expected to facilitate in monitoring and control of transportation networks. Nevertheless, these curves of MFD are not unique. Their shape and properties depend on few factors and some factors are briefly discussed such as the location of loop detectors, spatial distribution of links densities, traffic signal settings and traffic mode decomposition.

Then an application of MFD by monitoring of transportation networks is surveyed in Section 2.3. Monitoring through MFD assist traffic engineer by providing the current state of traffic congestion. From the literature, most works conducted used different sources of data, either coming from fixed sources or mobile sources or a combination of both. However, it is apparent that a low penetration rate of probe data in data fusion exercises can reduce the accuracy of MFD estimation. In conclusion, whereas data from mobile sensors are widely-available throughout the network, the information obtained from loop detectors remains vital for the effective traffic monitoring of urban areas and the reconstruction of macroscopic traffic flow models. The context is narrowed to operational or sparse-measurement MFD, which involves less number of sensors and corresponding measurements. Limitation of the existing literature include no automated procedure, the location of measurements are overlooked and no reliable accuracy assessment of sparse MFD. The motivation is to develop a generic framework for efficiently building sparse-measurement operational MFDs. This study will contribute in two ways; develops a sound methodology for the near-optimal sensor selection across a transport network and present a meticulous information-theoretic based framework for the efficient model selection and construction of sparse-measurement MFD. Further description of the methodology and results are discussed in Chapter 3.

Next, traffic flow control of transportation networks is reviewed in Section 2.4. Perimeter flow control policy is introduced to improve mobility in single-region homogeneous and multi-region heterogeneous networks. A perimeter flow control policy "meters" the input flow to the system and hold vehicles outside a protected network area, so as to maximise the throughput. Then different control approaches using the concept of MFD in the literature are briefly discussed such as bang-bang control, feedback control, MPC, robust control and switching control. Studies on perimeter flow control assume that a single input flow ordered by a perimeter control strategy should be equally distributed to a number of candidate junctions at the periphery of the network, i.e., without taking into account the different geometric characteristics of origin links, i.e., length, storage capacity, etc. Therefore, the motivation is to develop an integrated model of multi-gated perimeter flow control which take into account limited origin links storage capacity/geometric characteristics and the requirement of equity for drivers using different gates to enter a protected area. This study will contribute in determining optimal distribution of input flow values (or feasible entrance link green times) and providing equity for drivers using different gates to enter the protected network area. Further description of the methodology and results are discussed in Chapter 4.

Chapter 3

Traffic monitoring of transport networks

3.1 Introduction

This chapter presents an information-theoretic framework for the optimal selection of sensors across a traffic network. Sensor selection is of particular importance for the development of operational or sparse-measurement NFDs. A sparse-measurement NFD is less costly and it can be used for the efficient monitoring and perimeter control of congested urban areas. For the selection of sensors a generalised set covering integer programming problem (GIP) is developed. A measure of correlation between random variables, reflecting a variable of interest, is introduced as a "distance" metric to provide sufficient coverage and information accuracy. The Kullback-Leibler divergence is used to measure the dissimilarity between probability mass functions corresponding to different solutions of the GIP program. The performance of the proposed approach is assessed using empirical inductive loop-detector data spanning one week from a central business district with around sixty sensors. Results demonstrate that the obtained sparse-measurement rival models are able to preserve the shape and main features of the full-measurement traffic flow models e.g., flow capacity, critical occupancy (or density) and free-flow space-mean speed.

Concluding this chapter contributes in the state-of-the-art in two ways: (a) it provides a sound methodology for the selection of sensors across a transport network for the development of operational or sparse-measurement NFDs; and, (b) it presents a rigorous information-theoretic based framework for the efficient model selection. Notably, this framework can be adopted and used in other sensor location problems in location science.

The rest of the chapter is organised as follows. Section 3.2 introduces the sparse-measurement NFD, formulates the sensor selection problem, and introduces the Kullback-Liebler divergence for model selection. Section 3.3 presents a case study for a central business district using empirical data from fifty-eight sensors. The proposed information-theoretic framework for the optimal selection of sensors is then assessed and compared with two ad-hoc strategies. Section 3.4 summarises the main findings. Appendix A provides the required mathe-

matical background on information theory to assist the reader.

3.2 Sparse-measurement NFD and selection of sensors

3.2.1 Sparse-measurement NFD

A full-measurement diagram is constructed by flow-occupancy (or proxies of occupancy) measurements of n inductive-loop detectors (or other sensors) placed at appropriate network locations. A sparse-measurement diagram can be then constructed by selecting only a number of $k \leq n$ sensors. Clearly, different levels of network coverage in terms of selected sensors k (virtually percentage of monitored links) can provide different levels of accuracy. In principle, this is a combinatorial problem where the number of possible combinations is given by

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

Depending on the size of the network and number of sensors, checking these combinations would be overwhelming and practical impossible. As an example, the San Francisco network (financial district and south of market area) used in [1] includes around $n = 400$ links. If $k = 100$ links selected (25%) then the number of sparse-measurement NFDs to be examined is 2.24×10^{96}. Note that the total number of elementary particles in the universe is around 10^{80} (the Eddington number).

To overcome this difficulty, this work proposes to formulate the network coverage problem as a sensor ("facility") location problem. The idea here is to select a subset of links from a given candidate set, and place in each of these links a "sensor" that will provide flow-occupancy measurements to construct the sparse-measurement NFD. This is combinatorial optimisation problem and can be formulated as an integer programming problem where a 0-1 decision variable associated with selecting any given link for sensor placement, at a given cost.

3.2.2 Optimal selection of sensors

For the selection of sensors a set covering model is employed since number of sensors p to be selected is not known in advance, in contrast with $p-$ccenter, $p-$median models or maximal covering location problem. Moreover, a set covering model can provide the minimum number of sensors required together with the locations which ensure optimum level coverage. A formal definition of the set covering problem is as follows:

Definition 1 (Set Covering Problem)**.** *Let $\mathcal{U}$ be a finite set of cardinality n and let $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ be a family of subsets of $\mathcal{U}$, whose union equals the universe, i.e., $\mathcal{U} = \bigcup_{j=1}^{m} S_j$. Find a minimum-cardinality subfamily $\mathcal{C} \subseteq \mathcal{S}$ that covers the universe set $\mathcal{U}$, i.e., the union of all sets in $\mathcal{C}$ is $\mathcal{U}$.*

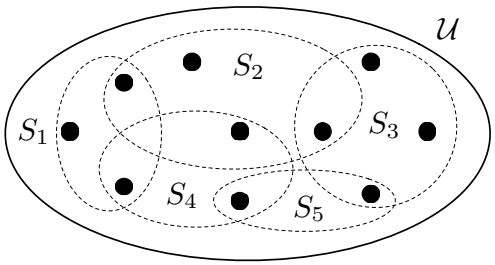


Figure 3.1: The set covering problem.

Clearly, such a cover exists if and only if the union of all sets in $\mathcal{S}$ is $\mathcal{U}$, and this is assumed for the rest of the chapter. For instance, a network with n sensors and universe $\mathcal{U} = \{1, 2, \ldots, n\}$ can be covered by $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$, where each $S_j = \{j\}$ is a singleton. Of course the cardinality of $\mathcal{S}$ in this case is n (i.e., maximum), given that all sensors selected. On the other hand, multiple sensors can be assigned to a single set S_j if they are correlated and provide the same amount of information or coverage, and thus the cardinality of $\mathcal{S}$ can be accordingly reduced (see Figure 3.1). The ultimate objective is to find a minimum-cardinality subfamily $\mathcal{C}$ of $\mathcal{S}$ that covers the universe set $\mathcal{U}$. The cardinality $k \leq m$ of $\mathcal{C}$ is free and will be specified by the optimisation. In contrast, in a maximal coverage or p-median problem an integer $k \leq m$ is also specified as input, and the goal is to select k subsets from $\mathcal{S}$ such that their union has the maximum cardinality.

The set covering problem can be extended to its weighted version, where each set $S_j \in \mathcal{S}$ is associated with a positive cost w_j and the objective is to find a cover with minimum total cost. It can also be formulated as a stochastic fixed-charge facility location problem in order to balance the trade-off between sensor allocation costs and information loss.

The standard integer programming (IP) problem formulation reads:

$$\begin{aligned} \text{(IP)} \ \min_{\boldsymbol{\chi}} \quad & f(\boldsymbol{\chi}) = \sum_{j=1}^{m} w_j \chi_j \\ \text{subject to:} \quad & \sum_{j=1}^{m} c_{ij} \chi_j \geq 1, \quad \forall\, i \in \mathcal{U}, i \in S_j, \\ & \chi_j \in \{0, 1\}, \quad \forall\, j \in \{1, \ldots, m\}, \end{aligned} \tag{3.1}$$

where $\mathcal{U}$ is the universe set, w_j is the fixed cost of assigning a sensor to set S_j, c_{ij} is a binary covering constant that takes the value 1 if element of $i \in \mathcal{U}$ is covered by set S_j (within "distance" δ) and the value 0 otherwise. Finally, a variable χ_j is introduced for every set S_j, with the intended meaning that $\chi_j = 1$ when S_j is selected, and $\chi_j = 0$ otherwise. The solution of this problem $\boldsymbol{\chi}^*$ provides the optimal sensor selection (selected sensors) and the minimum-cardinality of $\mathcal{C}$ where $k \triangleq \mathrm{card}(\mathcal{C}) = f(\boldsymbol{\chi}^*)$, provided $w_j = 1, \forall\, j$.

Now a Generic Integer Programming (GIP) formulation is introduced for the covering prob-

lem. The GIP is an enhanced version of (3.1) and involves a number of criteria reflecting features of the original full-measurement NFD that should be preserved in the developed sparse-measurement NFDs. The GIP is data-driven model under available flow-occupancy measurements of a number of sensors and the spatial distance of neighbour sensors. The GIP reads:

$$
\begin{aligned}
\text{(GIP)} \min_{\boldsymbol{\chi}} \quad & f(\boldsymbol{\chi}) = \sum_{j=1}^{m} w_j \chi_j \\
\text{subject to:} \quad & \sum_{j=1}^{m} c_{ij}^{of} \chi_j \geq 1, \quad \forall\, i \in \mathcal{U}, i \in S_j, && (3.2)\\
& \sum_{j=1}^{m} c_{ij}^{or} \chi_j \geq 1, \quad \forall\, i \in \mathcal{U}, i \in S_j, && (3.3)\\
& \sum_{j=1}^{m} c_{ij}^{fr} \chi_j \geq 1, \quad \forall\, i \in \mathcal{U}, i \in S_j, && (3.4)\\
& \chi_j \in \{0,1\}, \quad \forall\, j \in \{1,\ldots,m\},
\end{aligned}
$$

where c_{ij}^{of}, c_{ij}^{or} and c_{ij}^{fr} are binary covering constants which reflect different combinations of occupancy, flow and spatial distance. Particularly, c_{ij}^{of}, c_{ij}^{or} and c_{ij}^{fr} take value 1 if element of $i \in \mathcal{U}$ is covered by set S_j (within some predefined "distance") and the value 0 otherwise. The notion of "distance" is discussed later on, see (3.5)–(3.7). Equation (3.2) reflects information that could be acquired from occupancy and flow observations. Equation (3.3) reflects information that could be acquired from the combination of occupancy data and spatial distance of neighbour sensors. Equation (3.4) reflects information that could be acquired from the combination of flow data and spatial distance of neighbour sensors. Finally, a variable χ_j is introduced for every set S_j, with the intended meaning that $\chi_j = 1$ when S_j is selected, and $\chi_j = 0$ otherwise. The solution of this problem $\boldsymbol{\chi}^*$ provides the optimal sensor placement (selected sensors) and the minimum-cardinality of $\mathcal{C}$ where $k \triangleq \text{card}(\mathcal{C}) = f(\boldsymbol{\chi}^*)$, provided $w_j = 1, \forall\, j$.

Although the set covering problem is in general NP-hard to solve efficient approximation algorithms can be used to solve the problem in polynomial time. Problems with integer constraints can be usually approximately solved through some heuristic that neglects the integer constraints. In particular, a problem can be solved as a continuous problem and then use some ad-hoc method to round the fractional solution to integer. Heuristics often need to strengthen with more systematic procedures that can provide some assurance of an improved solution. The most well-known approaches include the greedy, the Linear Programming (LP) relaxation, the primal-dual, and the LP-rounding techniques.

To solve the GIP problem above the binary covering constants c_{ij}^{of}, c_{ij}^{or} and c_{ij}^{fr}, for all $i \in \mathcal{U}, i \in S_j$, must be specified. These matrices can be constructed if a notion of "distance" is appropriately defined. In the classic facility location or fixed-charge problems "distance" has the meaning of spatial distance. Here the spatial distance cannot be employed as a single metric but

can be combined with other information given that sensors or links located (far one each other) at arbitrary areas of the network can provide similar information and thus should be excluded. Another reason of not selecting the spatial distance as a single metric is that the spatiotemporal distribution of congestion in a network affects the NFD shape and its characteristics [36, 71]. [27] also demonstrated that signal control of a junction is affected by the traffic conditions even of relatively distant links for the network of Glasgow city (independent if the junction located centrally or at the network boundaries). A measure of correlation based on Pearson correlation or mutual information (see Appendix A) between random variables can be used as a "distance" metric to provide sufficient coverage and information accuracy. According to information theory rare events provide in average more information when observed, which is the case of sensors located at minor approaches but proximal to (surprisingly) less informative sensors at major approaches. The ultimate goal is to choose the minimum number of sensors keeping those that are most informative about unsensed locations, as explained in the sequel.

Consider a transport network with n sensors reporting flow and occupancy observations. Suppose that the time-occupancy (or flow) data in each sensor is described by a discrete random variable $X_i \in \mathcal{X}$ (the time index is omitted for clarity), $i = 1, 2, \ldots, n$ with $\mathcal{X} = \{0, 1, 2, \ldots, 100\}$ the finite set of occupancy observations (0-100%). The main idea here is to look for the correlation of all pairs (X_i, X_j) for all $i, j = 1, 2, \ldots, n$, see (A.6) and (A.8). High correlated random variables (with $\varrho \approx 1$ or $\rho \approx 1$) provide on average the same information (their expected value is the information entropy (A.1)), and thus their measurements contribute in the same way in the construction of the NFD, provided more or less similar flow observations. Therefore, it would be desirable the GIP problem to exclude a number of those detectors providing similar coverage. On the other hand, low correlated or independent random variables (with $\varrho \approx 0$ or $\rho \approx 0$) provide more information and their measurements are important for the construction of the NFD. It is thus desirable, the GIP problem to include those detectors in the final solution. Finally, the spatial distance is also take into account by considering neighbours covered within radius R (in meters or kilometers). In this way, occupancy (or flow) information and distance between sensors are compromised to make the formulation reasonable.

With these observations in mind a natural selection of "distance" for our problem will be: δ^o and δ^f present a threshold term that define relevant occupancy correlation and flow correlation respectively. Mathematically, $\delta^o = \varrho$ or $\delta^o = \rho$ and $\delta^f = \varrho$ or $\delta^f = \rho$ while R presents the spatial distance threshold (radius). For given δ^o and δ^f, the binary covering matrix $\mathbf{C} = [c_{ij}]$ is given by three different covering constants as follows:

$$c_{ij}^{of} = \begin{cases} 1, & \text{if } d_{ij}^o \geq \delta^o \text{ and } d_{ij}^f \geq \delta^f \\ 0, & \text{otherwise,} \end{cases}, \qquad i, j = 1, 2, \ldots, n, \tag{3.5}$$

Table 3.1: Candidate covering models derived from the GIP.

Model	Properties (Subject to:)	Binary covering constant c_{ij}	Criteria		
			δ^o	δ^f	R
Model 1	(3.4)	(3.7) or (3.6) without $r_{ij} \leq R$	-	✓	-
Model 2	(3.2)	(3.5)	✓	✓	-
Model 3	(3.4)	(3.7)	-	✓	✓
Model 4	(3.3)–(3.4)	(3.6)–(3.7)	✓	✓	✓

$$c_{ij}^{or} = \begin{cases} 1, & \text{if } d_{ij}^o \geq \delta^o \text{ and } r_{ij} \leq R, \\ 0, & \text{otherwise,} \end{cases} \qquad i, j = 1, 2, \ldots, n, \tag{3.6}$$

$$c_{ij}^{fr} = \begin{cases} 1, & \text{if } d_{ij}^f \geq \delta^f \text{ and } r_{ij} \leq R, \\ 0, & \text{otherwise,} \end{cases} \qquad i, j = 1, 2, \ldots, n, \tag{3.7}$$

The binary variable c_{ij}^{of} in (3.5) (see (3.2) in GIP) ensures that information from both occupancy and flow observations is used. The binary variable c_{ij}^{or} in (3.6) (see (3.3) in GIP) ensures that information from occupancy observations and the spatial distance is used, where r_{ij} is the radius between sensors i and j. Similarly the binary variable c_{ij}^{fr} in (3.7) (see (3.4) in GIP) ensures that information from flow observations and the spatial distance is used. For the rest of the chapter $d_{ij}^o = \rho(X_i, Y_j)$, $d_{ij}^f = \rho(X_i, Y_j)$ is assumed. Table 3.1 summarises the properties of a number of candidate models that could be derived from the GIP and used in this thesis for reconstructing the sparse-measurement NFD.

Clearly if $\delta^o = \delta^f \equiv 1$ and R is big enough to cover the whole network, then $\mathbf{C} = \mathbf{I}_n$ and the IP optimal solution is $f(\boldsymbol{\chi}^*) = n$, i.e., all sensors are selected $\chi_j = 1, \forall\, j$ (full-measurement NFD). If $\delta^o < 1$ or $\delta^f < 1$ and R is small then the optimisation will exclude a number of sensors and a sparse-measurement NFD can be constructed from the selected sensors. Obviously, the accuracy of the NFD approximation will be reduced (e.g., its shape will change) as $\{\delta^o, \delta^f, R\}$ decreased, but the optimal value of the objective function in GIP corresponding to the cost will be improved. This procedure reduces significantly the number of sensor locations that need to be checked, thus speeding up computation of the IP. Moreover, it provides an upper-bound on the value of the optimal selection (for $\delta^o = \delta^f = 1$ and R big enough), which can be used as a worst-case (best approximation) bound by other exact (e.g., branch-and-bound) or heuristic approaches.

To check the accuracy of the NFD approximation for a particular criteria (i.e., for a particular GIP model) its sparse-measurement probability mass function are calculated. Then the sparse-measurement candidate probability mass function is compared with an empirical (ground truth)

reference probability mass function reflecting the full-measurement NFD. In this way one can measure how much information is lost and how good is the NFD approximation.

3.2.3 Kullback-Leibler divergence and model selection

To check the accuracy of the NFD approximation, this work employs the information-theoretic relative entropy (A.2), known as the Kullback-Leibler divergence (KL-divergence), to measure the dissimilarity or "distance" between probability mass functions corresponding to different models obtained from the solution of the GIP problem for different criteria. Entropy (A.1) reflects the average information included in our data set and gives a theoretical lower bound on the number of bits needed to encode a random variable X or its probability mass function $p_X(x)$. The relative entropy or KL-divergence reflects the average loss of using another code (or model $q(x)$) to encode a random variable X or its probability mass function $p(x)$.

Certainly, the interest is in probability mass functions or models that preserve the most information from the original data source (i.e., from the empirical probability mass function). The relative entropy can also be interpreted as the information gain achieved about X if p can be used instead of q. Under certain regulatory conditions, KL-divergence is a monotonically decreasing function with information gain, while it is zero if and only if two distributions are equal. Therefore KL-divergence minimisation leads to information loss minimisation. On the other hand, the optimal value of the objective function in GIP is improved with information loss as δ^o or δ^f is decreased while KL-divergence is increased. Note that KL-divergence is unbounded from above, see the discussion in Section 3.3.5. Concluding, efficient model selection (and sparse-measurement NFD approximation) is a trade-off between the KL-divergence and the cost of the objective function in GIP.

3.3 Results

3.3.1 Data description and setup

Flow-occupancy experimental data (1.5-min samples) from 58 inductive-loop mid-block detectors shown in Figure 3.2 and spanning one week in June 2006 [60], were available for the testing of the proposed data inference and model selection scheme. These data sets were collected in a field evaluation of the TUC/HYBRID signal control strategy and the commercial semi-real-time strategy TASS developed by Siemens in the central business district (CBD) of Chania, Greece [60]. The CBD includes about 24 closely spaced signalised junctions and 71 links with varying lengths. The CBD is congested during the weekday peaks (especially in the summer due to tourism) and sometimes lead to partial gridlock situations. Traffic conditions in the network are quite different, even among weekdays, due to differences in shop opening times. In [7] showed that the CBD of Chania exhibits a network-wide fundamental diagram that is reproduced under

Table 3.2: Capacity flow and critical occupancy for one week.

Day	Capacity flow (veh/cycle)	Range of critical occupancy (%)
Monday	750	16-23
Tuesday	700	18-40
Wednesday	700	18-23
Thursday	740	20-25
Friday	750	20-25
Saturday	700	20
Sunday	640	15

different traffic conditions (different days) but its shape and critical occupancy depend on the applied semi-real-time signal control and the distribution of congestion in the network.

3.3.2 Full-measurement NFD

Figure 3.3 and 3.4 (reproduced from [7]) depict the full-measurement NFD (circulating flow versus occupancy obtained from 58 sensors) for one week (Monday to Sunday). Each measurement point on the fundamental diagram corresponds to 1.5 min samples of flow and occupancy (%) and the data were aggregated to 6 min all-lane data by averaging the lane occupancies and summation of flows. These figure confirm the existence of an NFD for the CBD of Chania over one week. The shape observed is similar across different days and the congested regime of the NFD is only partly formed, indicating that throughput (flow capacity) is maintained in the network. The maximum flow and critical occupancy values are given in Table 3.2. As can be seen maximum flow is observed in the range 600-750 veh/cycle for different critical occupancies in different days.

The following observation reproduced from [7]. On Monday and Wednesday, flow capacity is observed to have similar range of critical occupancies and only part of the congested regime appears in the diagrams (see Figure 3.3(a) and 3.3(c)). On Tuesday, Thursday and Friday where shops are open in the evening, flow capacity is observed over a range of occupancies. The NFDs of Tuesday and Thursday in Figures 3.3(b) and 3.3(d) indicating crowded traffic condition (see occupancy level exceed 30%) but maintaining throughput thanks to efficient signal control. The NFD of Friday in Figure 3.3(e) visits the congested regime. On Saturday and Sunday in Figure 3.4 only part of NFD is formed. This is attributed to low demand and the shops are closed especially on Sunday.

Figure 3.5–3.8 illustrate flow and occupancy time series spanning one week. Regarding the flow time series, the onset and offset of congestion is pronounced for all days at: i) 05:00–17:00,

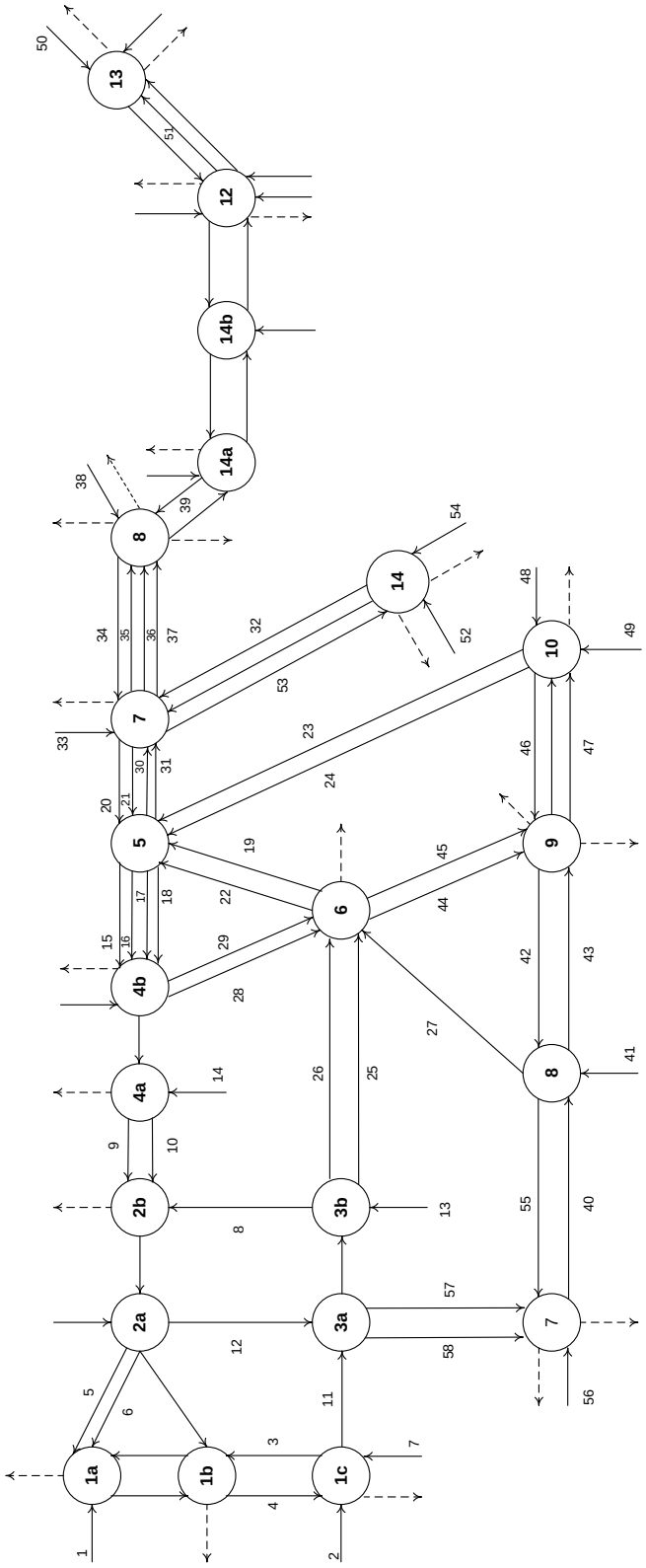
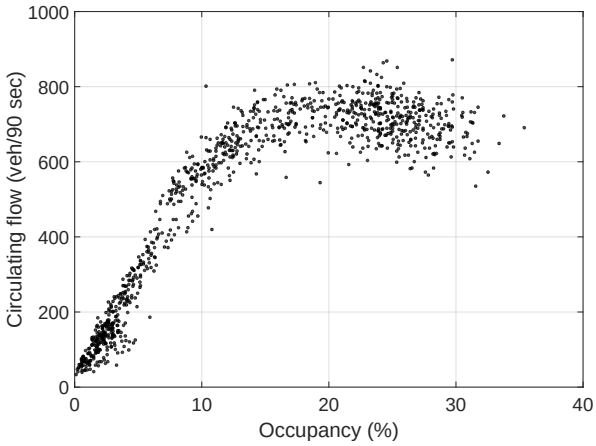


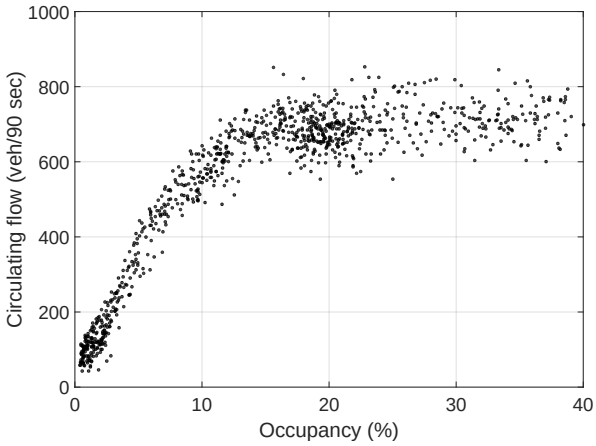
Figure 3.2: Schematic map of Chania, Greece [60].

and ii) 17:00–05:00. During the AM peak there is a significant rise of flow until approximately at 09:00 before the flow is maintained at its throughput level. However, in the weekend and especially on Sunday, there is a drop in flow capacity. This is likely due to the closed shops on Sunday. The PM peak corresponds to the evening rush hour. There is a significant drop of flow approximately around 23:00. This trend is observed in all days except Sunday where flow drops only after 01:00. Monday indicates high occupancies, 36% is observed at 11.00. Particularly on Tuesday, Thursday and Friday in Figure 3.6(a), 3.7(a), and 3.7(b), flow capacity is observed over a range of critical occupancies (18:00-22:00), where shops are open and saturated traffic condition are formed. In the weekend (Saturday and Sunday) occupancies remain below 30%.

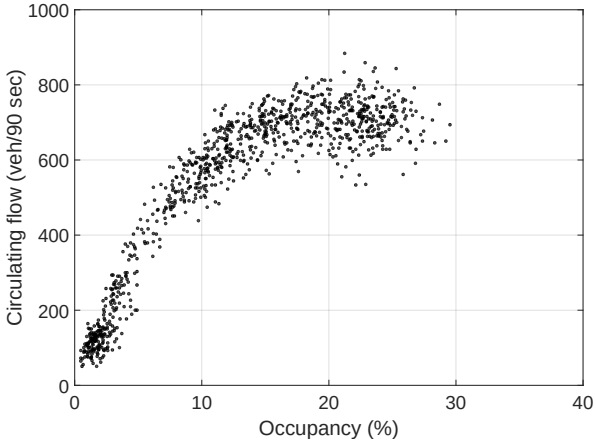
Figure 3.3 and 3.4 depict the full-measurement NFD, where flow corresponds to the average circulating flow in the network for all links. For assessing different sparse-measurement NFDs (involving different number of sensors/links, and thus different flows) in the sequel and having comparable results, Figure 3.3 and 3.4 are reproduced where flow is calculated in vehicles per hour per lane (or sensor/link). These NFDs are depicted in Figures 3.9–3.10 while Table 3.3 presents the corresponding maximum flow and critical occupancy values.



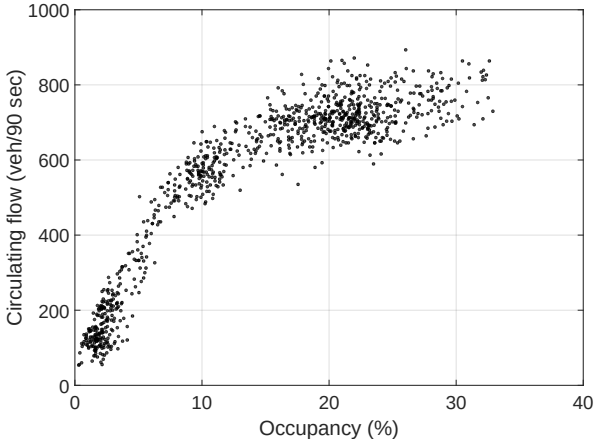
(a): Monday NFD



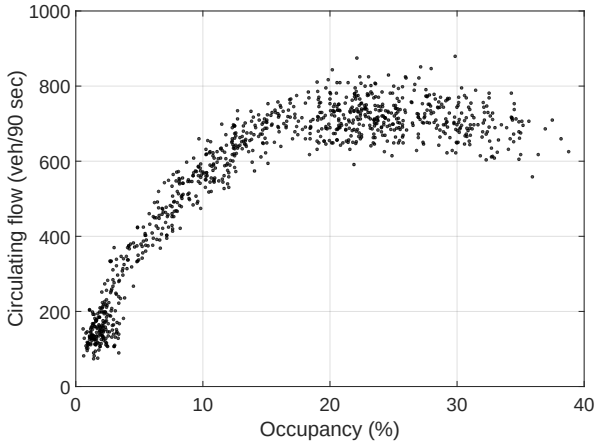
(b): Tuesday NFD



(c): Wednesday NFD

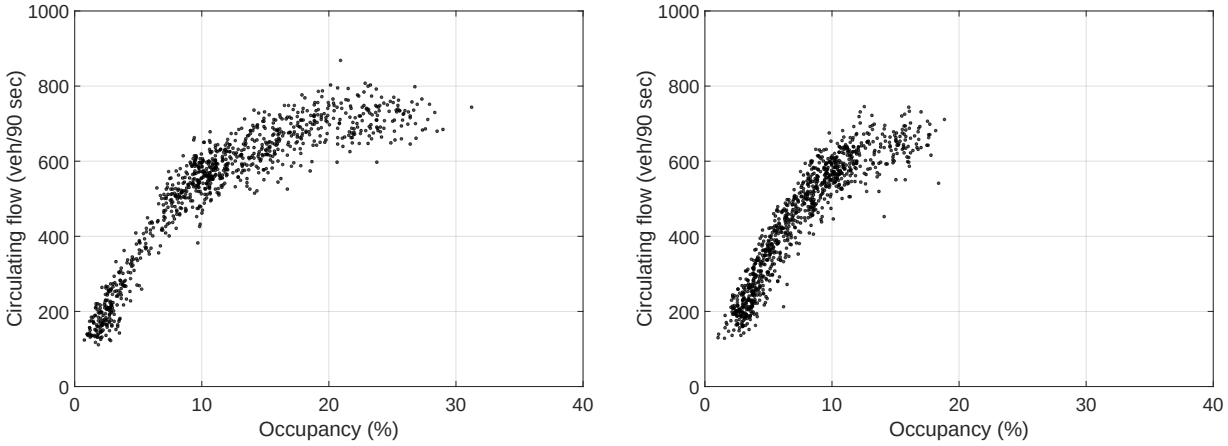


(d): Thursday NFD



(e) Friday NFD

Figure 3.3: Network-wide NFDs for Monday to Friday.



(a) Saturday NFD (b): Sunday NFD

Figure 3.4: Network-wide NFDs for Saturday and Sunday.

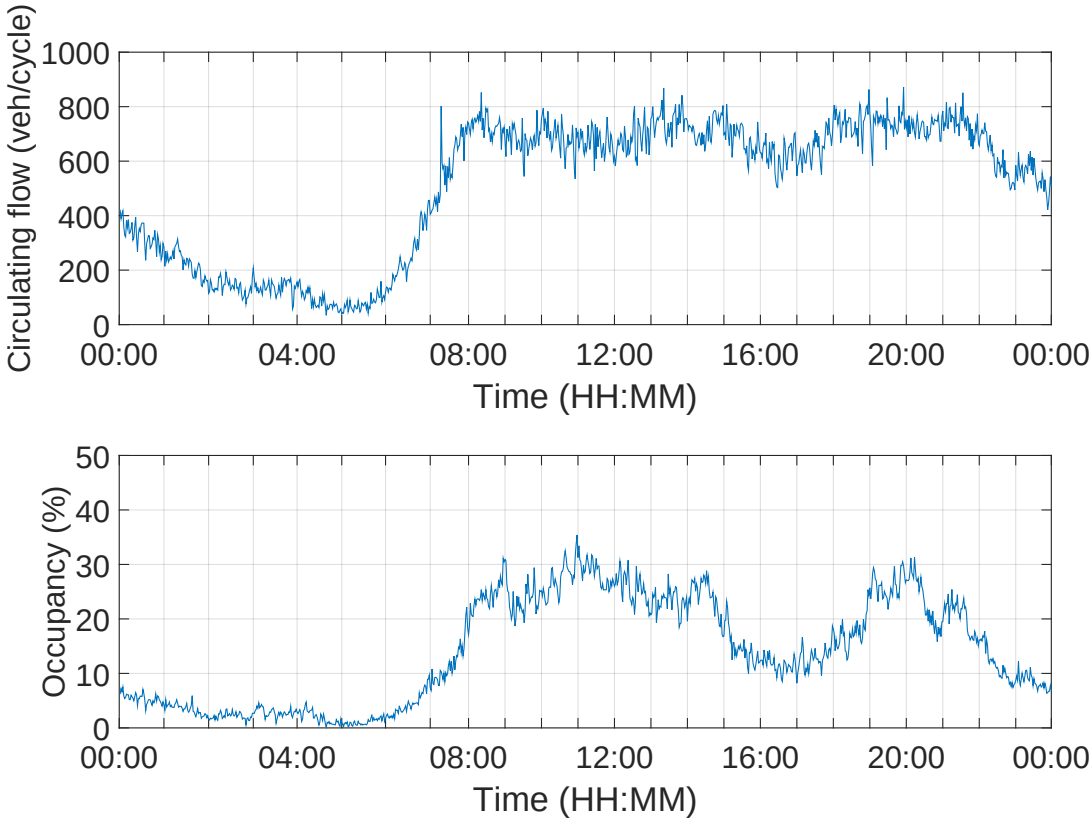
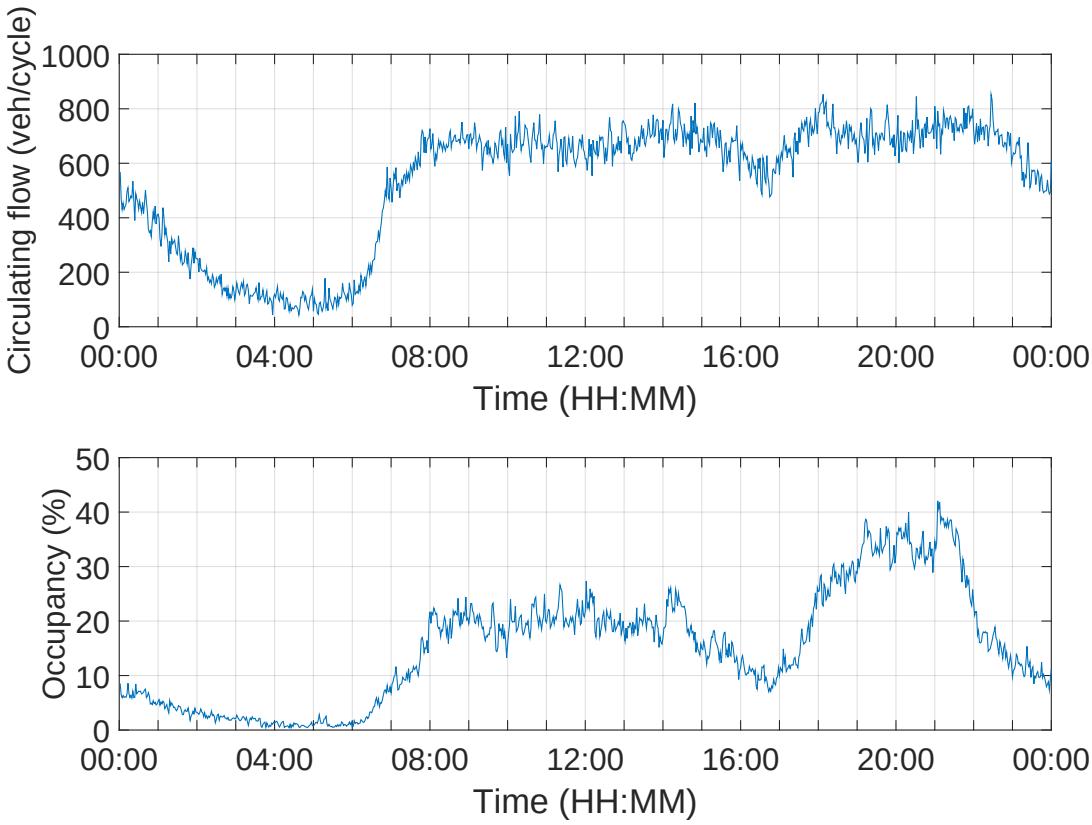
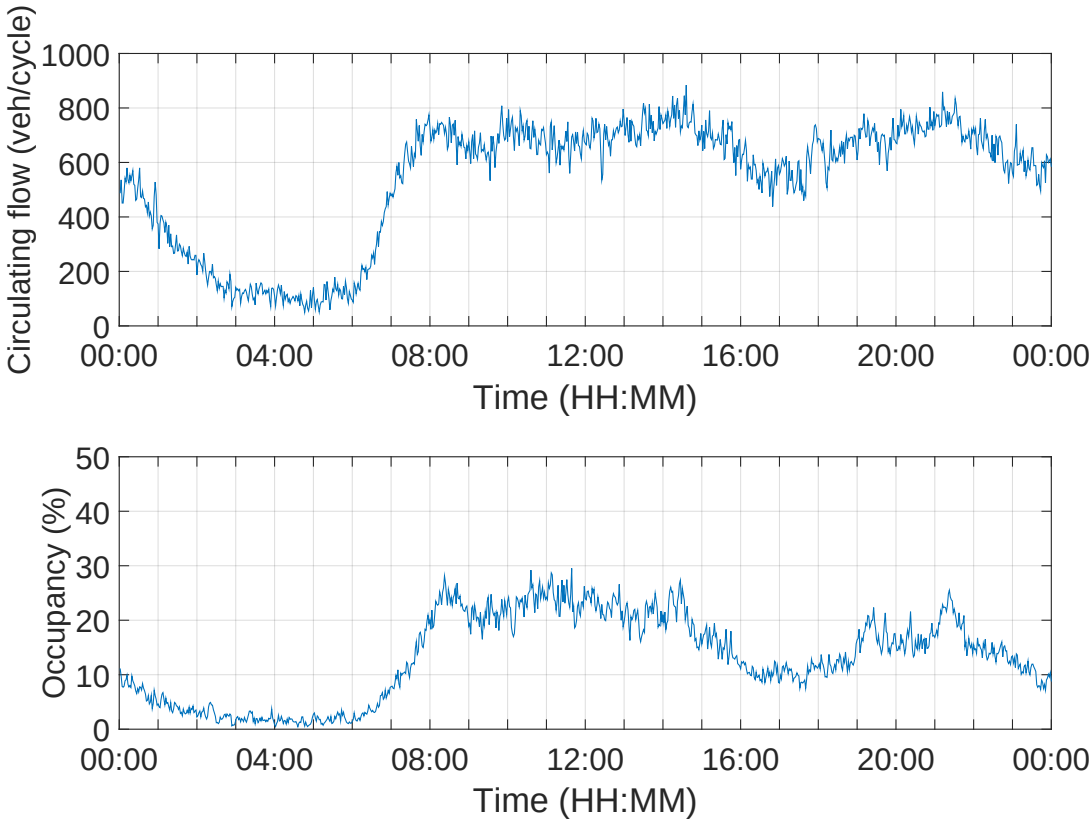


Figure 3.5: Flow and occupancy time series of Monday.

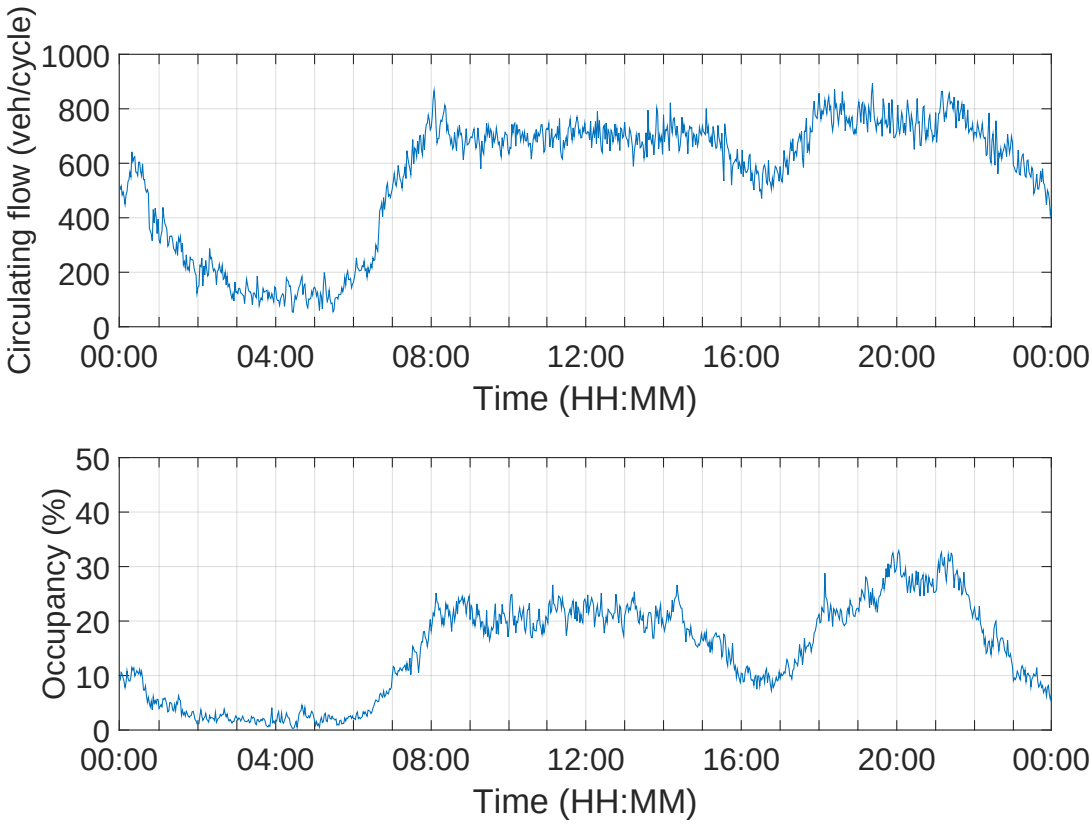


(a): Tuesday, June 6, 2006

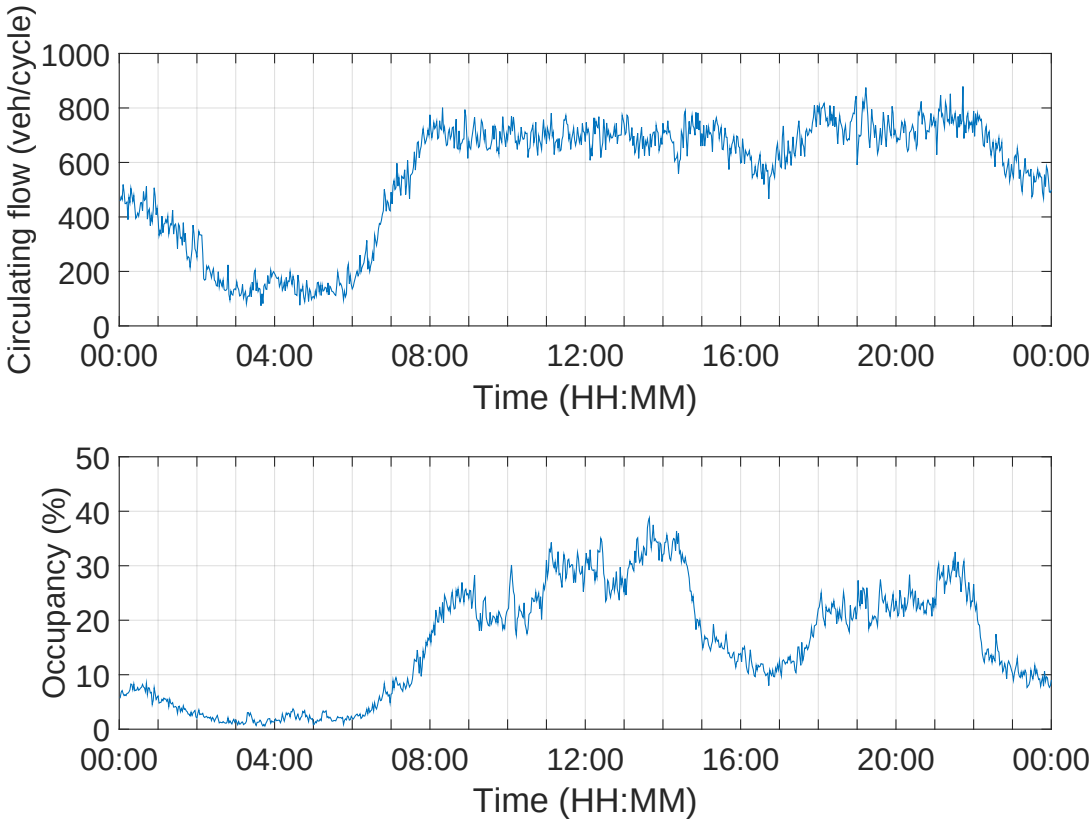


(b): Wednesday, June 7, 2006

Figure 3.6: Flow and occupancy time series of Tuesday and Wednesday.

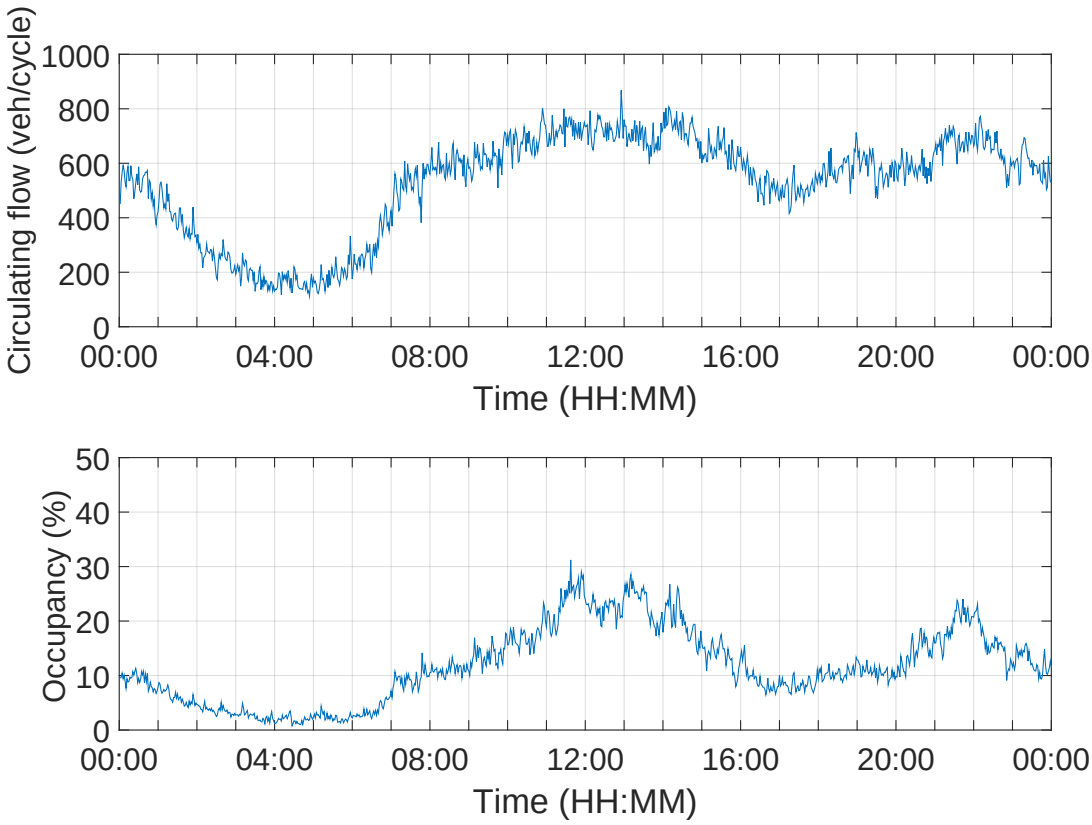


(a): Thursday, June 8, 2006

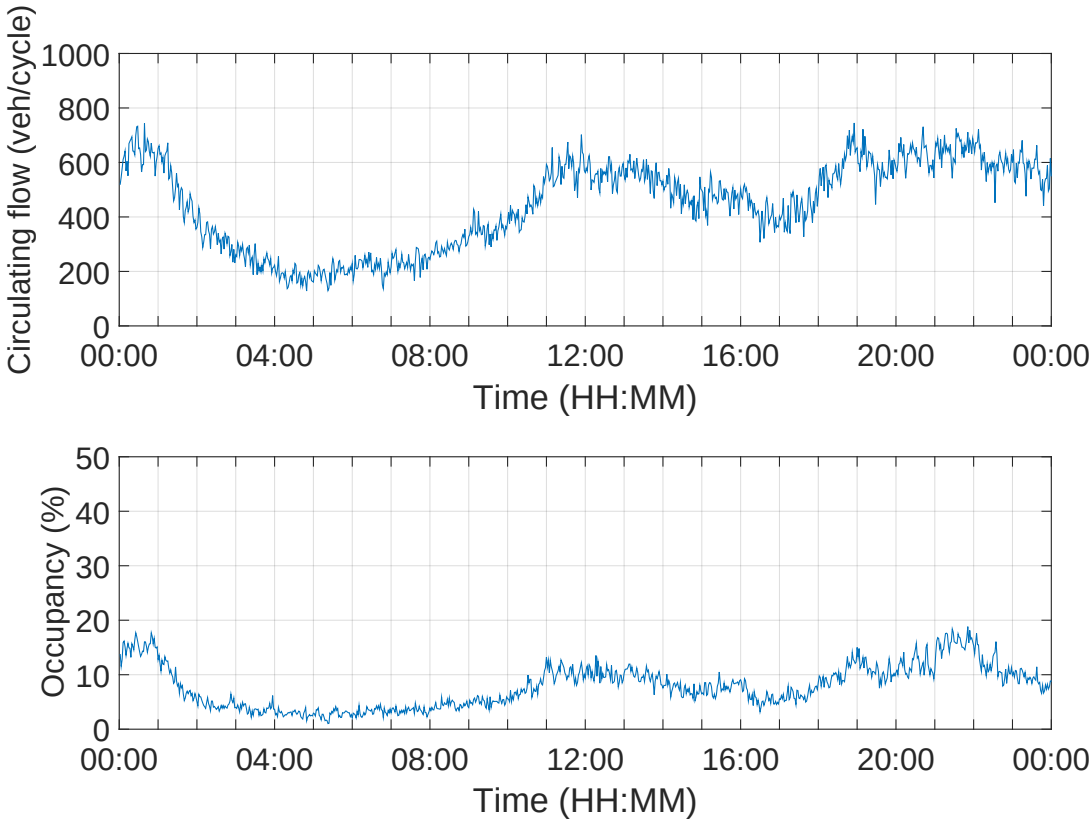


(b): Friday, June 9, 2006

Figure 3.7: Flow and occupancy time series of Thursday and Friday.



(a): Saturday, June 10, 2006

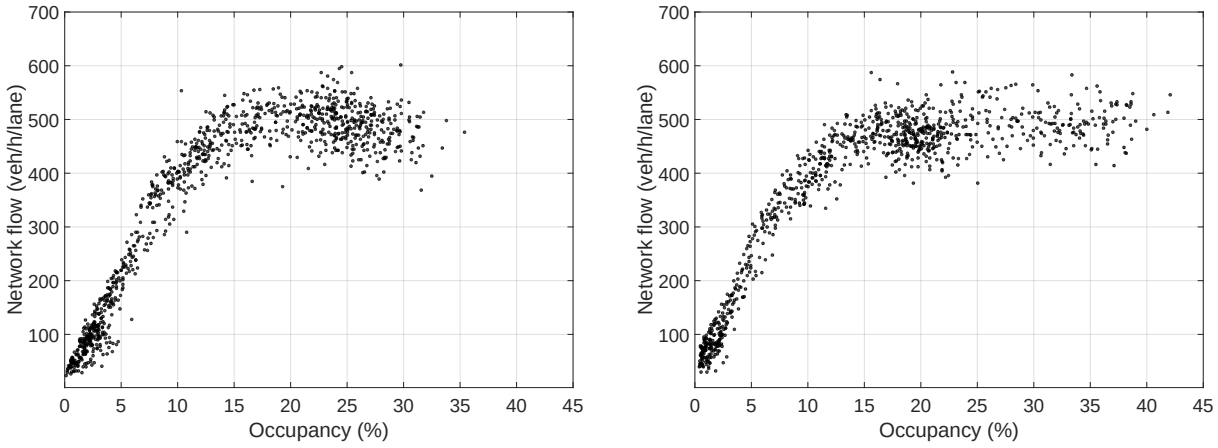


(b): Sunday, June 11, 2006

Figure 3.8: Flow and occupancy time series of Saturday and Sunday.

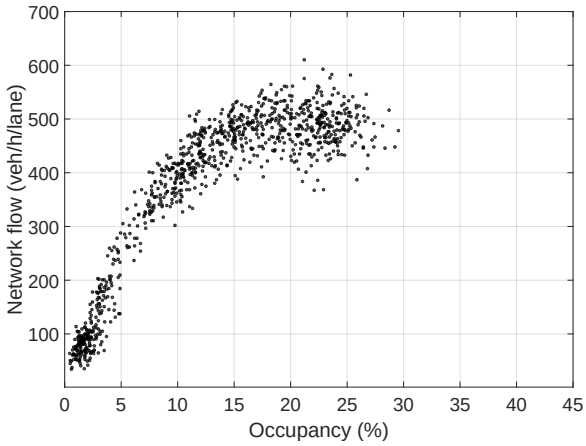
Table 3.3: Capacity flow and critical occupancy of full-measurement NFD for one week.

Day	Capacity flow (veh/h/lane)	Range of critical occupancy (%)
Monday	500	16-23
Tuesday	500	18-40
Wednesday	500	16-21
Thursday	520	20-25
Friday	500	20-25
Saturday	500	20
Sunday	470	15

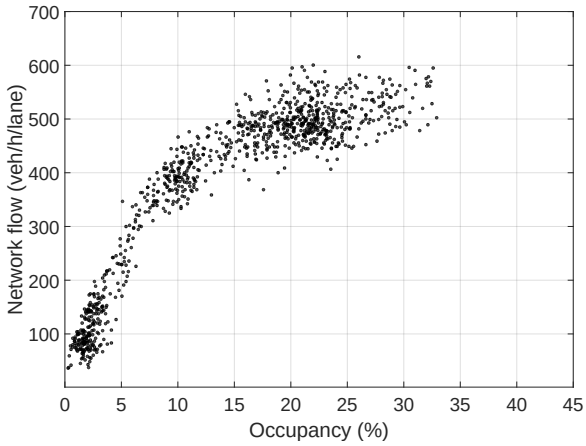


(a): Monday NFD (b): Tuesday NFD

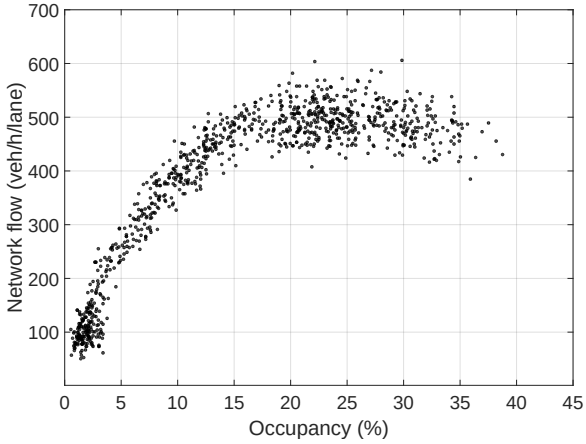
Figure 3.9: Network-wide full-measurement NFDs for Monday and Tuesday.



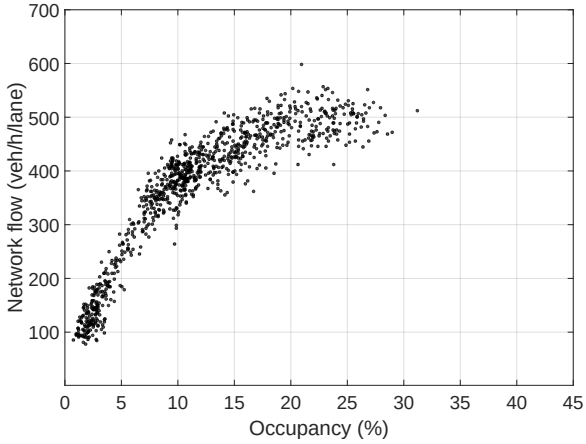
(a): Wednesday NFD



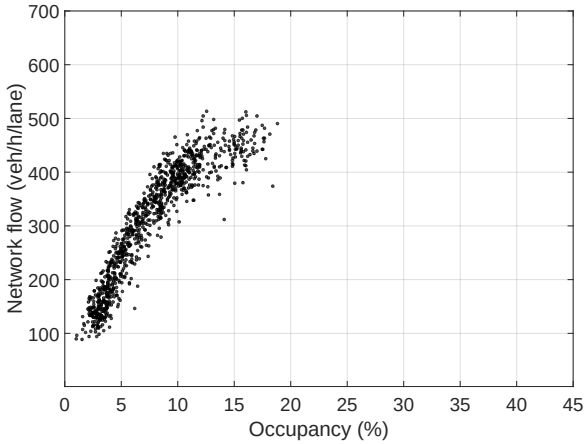
(b): Thursday NFD



(c) Friday NFD



(d): Saturday NFD



(e) Sunday NFD

Figure 3.10: Network-wide full-measurement NFDs for Wednesday to Sunday.

3.3.3 Optimal sensor selection

In this section, the full-measurement NFD for each day is approximated by a number of models (and corresponding sparse-measurement NFDs) obtained from the solution of the GIP problem. Model 1 is examined first for different $\delta^f \in \{0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 1\}$ due to its basic quantity and properties compared to other model mentioned in Table 3.1, Section 3.2.2. These δ^f values were selected to investigate the effectiveness and accuracy of approximation of the developed framework when highly correlated data (in descending order) are excluded from the GIP formulation (as reflected in $\mathbf{C}$ matrix). For each GIP solution (δ^f value) the minimum-cardinality subfamily $\mathcal{C} \subseteq \mathcal{S}$ is obtained and then used to calculate the corresponding probability mass function q^{δ^f}.

Consider a random vector $\mathbf{X}$ of random variables $X_1, X_2, \ldots, X_k$ (the time index is omitted for clarity) defined in the finite set $\mathcal{X} \in \{0, 1, 2, \ldots, 100\}$ (discrete occupancy observations 0-100%), with $k \triangleq \text{card}(\mathcal{C}) = f(\boldsymbol{\chi}^*)$. The value $p_{\mathbf{X}}(x) = \mathbf{P}(X_k = x, \text{for all } k)$ is the probability that the random vector $\mathbf{X}$ takes the value x. Now for a given time-occupancy data set the probability mass function $p(x)$ is calculated as

$$\mathbf{P}(X_k = x, \text{for all } k) = \frac{\phi_x + 1}{N + \sum_{x=0}^{N-1} \phi_x} \tag{3.8}$$

with $N \triangleq \text{card}(\mathcal{X})$ and ϕ_x the frequency of observation x in the dataset. Equation (3.8) is a simple smoothing technique to obtain a satisfactory probabilistic model in the presence of data sparsity (e.g., if there are many events x unobserved with $\phi_x = 0$). Note that entropy and KL-divergence require $p(x), q(x) > 0$ and probability mass functions that sum to one. Given the probability mass functions $q^{\delta^f}(x)$ for each GIP solution, i.e., for each $\delta^f \in \{0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 1\}$, entropy and KL-divergence $\Delta(p \,\|\, q^{\delta^f})$ are then calculated to investigate the accuracy of different models obtained. Here $p(x)$ is the empirical (ground truth) reference probability mass function reflecting the full-measurement NFD.

Integer solutions

Table 3.4–3.7 display the obtained results from the solution of the GIP problem (Model 1) for the one week data set across different $\delta^f \in \{0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 1\}$. As can be seen, the optimal value of the GIP solution as reflected from the cardinality of $\mathcal{C}$ is improved (i.e., less sensors selected) as δ^f decreases. On the other hand, KL-divergence is increased as δ^f decreases reflecting the information loss under data sparsity. An important observation is that the optimal value of the GIP solution is different from day to day for the same δ^f. For instance, $d^f \geq 0.7$ for Monday results 27 sensors (around 47%) while the same δ^f for Friday results 36 sensors (around 62%). Note that each GIP solution in Table 3.4–3.7 except the optimal cost of

sensor allocation (cardinality of $\mathcal{C}$) provides the subset of links (numbers in brackets) selected for sensor selection.

Probability mass functions

Figures 3.11–3.14 display the obtained probability mass functions (models) for Monday to Sunday and different $\delta^f \in \{0.65, 0.8, 0.9, 1\}$. The probability mass function for $\delta^f = 1$ (58 links) corresponds to the full-measurement NFD shown in Figures 3.9–3.10. As a first remark all models indicate a geometric distribution with $\mathbf{P}(X = x) = (1-p)^x\, p$, $x = 0, 1, 2, 3, \ldots$. This is particular useful when the parameter p of different δ^f models is estimated (maximum likelihood) from given occupancy observations and for calculating the relative entropy.

Figures 3.15–3.18 display comparison between probability mass functions of full measurements ($d^f \geq 1.0$) and its geometric distribution. As can be seen the geometric distribution closely follows the trend of the real distribution. The value of parameter p and its estimation $\hat{p}$ is given in Table 3.4–3.7 for all days across one week. This is beneficial where one can use the geometric distribution to approximate the probability mass functions, particularly in the absence of data or when sensors are malfunction and unable to provide reliable data.

To further explain this, the probability mass functions (pmf) on Friday is illustrated and some comparison is performed. Figure 3.19(a) illustrates the pmf for $d^f \geq 0.65$ and the full-measurement model for $d^f \geq 1.0$. The difference between the two distributions is shown in Figure 3.19(b). Figure 3.19(c) illustrates the pmf for $d^f \geq 0.55$ and $d^f \geq 1.0$. The difference between the two distributions is displayed in Figure 3.19(d). As can be seen the grey area is larger as δ^f decreases from 0.65 to 0.55 (see Figure 3.19(b) and 3.19(d)), reflecting information loss when δ^f decreases. This area illustrates the KL-divergence between the two distributions, which can be used to measure the dissimilarity between probability mass functions corresponding to different models.

KL-Divergence

Figure 3.20–3.21 depict the optimal cost of the IP problem (Model 1) versus KL-divergence for different δ^f over one week. In these plots, the primary vertical axis on the left is used for the optimal cost of the GIP problem, whereas the secondary vertical axis on the right side is for the KL-divergence. These plots indicate the trade-off between sensor allocation costs and information loss or information gain depending on how KL-divergence is interpreted. Figures 3.20(a) and 3.21(e) suggest that GIP solutions for any $\delta^f \geq 0.75$ are acceptable on Monday and Sunday, respectively. Correspondingly, at $\delta^f = 0.75$, number of 39 sensors for Monday and 46 sensors for Sunday can be selected for the construction of the sparse measurement NFD, since they provide the best trade-off between sensor allocation costs and information loss. Meanwhile, Figures 3.20(b) and 3.21(a–d) suggest that GIP solutions for any $\delta^f \geq 0.7$ are acceptable on

Tuesday to Saturday. At $\delta^f = 0.7$, selection of 39, 38, 37, 36 and 44 sensors for Tuesday to Saturday, respectively, is suggested for the construction of the sparse-measurement NFD.

Figure 3.22 presents collective results for all data sets spanning one week. Figure 3.22(a) gives the entropy for each day if 58 sensors selected ($\delta^f = 1$), the theoretical lower bound on the number of bits needed to encode the probability mass function of each day. As can be seen, Monday and Wednesday in which market and shops are closed in the evening indicate more or less the same lower bound. The same observation holds for Tuesday and Thursday in which the market and shops are open in the evening. These two days are more busy compared to Monday and Wednesday and thus they need more bits of information to encode their models. Friday is usually congested and indicates the higher lower bound among all days. In the summer most of traffic in the weekend is directed outside of the city particularly in Sunday, which indicates the lowest lower bound among all days. Figure 3.22(b) depicts the KL-divergence trajectories in function of δ^f for seven days. The KL-divergence value can be used as performance measure which quantify approximation error between the full measurement and the sparse (partial) measurement of each value of δ^f. This graph confirms that the KL-divergence is monotonically decreasing with respect to δ^f, it is nonnegative and zero if and only if two distributions are equal (case of $\delta^f = 1$). For $\delta^f = 0.6$ on Monday KL-divergence is decreased (compared to $\delta^f = 0.65$) due to the discrete nature of the pmf and data sparsity.

Sparse approximation

Figure 3.23–3.26 depict the sparse-measurement NFDs (circle markers in colour) obtained for selected models (δ^f values, see selected pmfs in Figures 3.11–3.14) when contrasted with the empirical full-measurement/reference or ground truth NFD (filled circle markers in black). By visual inspection, when δ^f increases and approaches 1.0, the approximation (circle markers in colour) is improved. As can be seen models with $\delta^f = 0.8$ and $\delta^f = 0.9$ provide excellent approximation of the full-measurement NFD. It can be also seen that all models preserve the shape and main features of the operational full-measurement diagram. These results underline the superiority of the proposed information-theoretic framework and KL-divergence in model selection.

Table 3.4: Solution of the set covering problem, entropy, KL-divergence and parameter of geometric distribution for Monday and Tuesday and different δ^f values.

Day		"distance" ($d^f \geq \delta^f$)							
		$d^f \geq 0.6$	$d^f \geq 0.65$	$d^f \geq 0.7$	$d^f \geq 0.75$	$d^f \geq 0.8$	$d^f \geq 0.85$	$d^f \geq 0.9$	$d^f \geq 1.0$
Monday June 5, 2006	Total # of selected links	13	20	27	39	48	52	56	58
	$\mathcal{C}_{\text{IP}}$	{4, 8, 14, 17, 23, 24, 25, 30, 33, 37, 49, 52, 57}	{2, 3, 4, 7, 8, 13, 14, 17, 19, 23, 24, 30, 32, 33, 37, 38, 45, 52, 56, 57}	{2, 3, 4, 7, 8, 11, 12, 13, 14, 15, 17, 19, 23, 24, 30, 32, 33, 37, 38, 39, 44, 48, 50, 52, 56, 57, 58}	{1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 15, 17, 19, 21, 23, 24, 25, 27, 29, 30, 31, 32, 33, 35, 37, 38, 40, 44, 47, 48, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15 ,16, 18, 19, 21, 23 , 24, 26, 27, 28, 30, 31, 32, 33, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 18, 19, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, …, 58}
	KL-div (bits)	0.038	0.038	0.022	0.019	0.0042	0.0015	0.00025	0
	Entropy (bits)	5.01	4.91	4.93	5.02	5.06	5.10	5.16	5.16
	$(p, \hat{p})$	$(0.209, 0.066)$	$(0.209, 0.069)$	$(0.188, 0.072)$	$(0.181, 0.068)$	$(0.173, 0.068)$	$(0.168, 0.068)$	$(0.161, 0.066)$	$(0.158, 0.066)$
Tuesday June 6, 2006	Total # of selected links	15	27	39	46	49	54	57	58
	$\mathcal{C}_{\text{IP}}$	{2, 3, 4, 8, 13, 14, 17, 23, 24, 30, 36, 38, 49, 52, 58}	{2, 3, 4, 7, 8, 12, 13, 14, 15, 17, 23, 24, 25, 30, 31, 32, 33, 37, 38, 43, 44, 48, 49, 52, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 19, 23, 24, 26, 27, 29, 30, 32, 33, 35, 37, 38, 40, 41, 44, 47, 48, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 19, 21, 23, 24, 26, 27, 29, 30, 31, 32, 33, 35, 36, 37, 38, 40, 41, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 19, 21, 23, 24, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, …, 58}
	KL-div (bits)	0.027	0.025	0.016	0.0077	0.0045	0.0011	0.000085	0
	Entropy (bits)	5.04	5.13	5.11	5.17	5.17	5.21	5.27	5.27
	$(p, \hat{p})$	$(0.185, 0.067)$	$(0.175, 0.063)$	$(0.165, 0.065)$	$(0.159, 0.063)$	$(0.153, 0.064)$	$(0.148, 0.064)$	$(0.143, 0.062)$	$(0.142, 0.061)$

Table 3.5: Solution of the set covering problem, entropy, KL-divergence and parameter of geometric distribution for Wednesday and Thursday and different δ^f values.

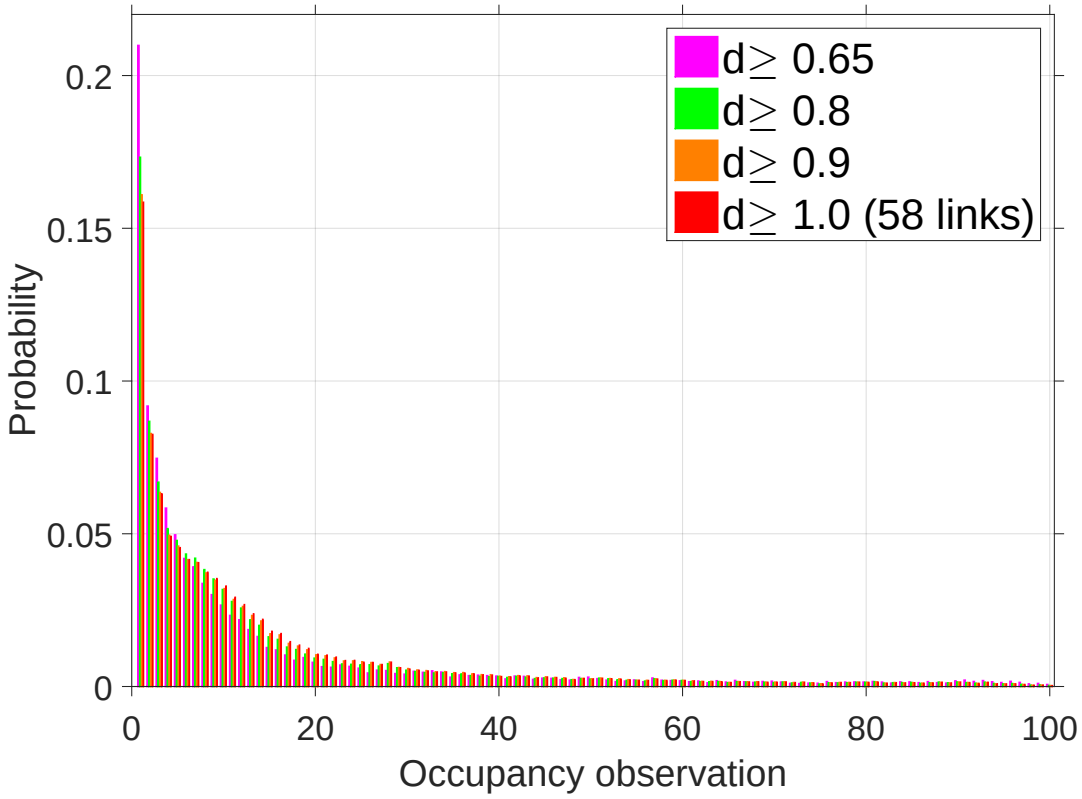
Day		"distance" ($d^f \geq \delta^f$)							
		$d^f \geq 0.6$	$d^f \geq 0.65$	$d^f \geq 0.7$	$d^f \geq 0.75$	$d^f \geq 0.8$	$d^f \geq 0.85$	$d^f \geq 0.9$	$d^f \geq 1.0$
Wednesday June 7, 2006	Total # of selected links	21	27	38	45	51	53	57	58
	$\mathcal{C}_{\text{IP}}$	{2, 3, 4, 7, 8, 11, 13, 14, 15, 17, 23, 24, 30, 32, 33, 37, 38, 44, 52, 56, 57}	{1, 2, 3, 4, 7, 8, 11, 13, 14, 15, 17, 23, 24, 25, 28, 30, 32, 33, 37, 38, 44, 48, 49, 52, 53, 56, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 19, 23, 24, 27, 29, 30, 32, 33, 35, 37, 38, 43, 44, 45, 47, 48, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 19, 21, 23, 24, 25, 27, 29, 30, 32, 33, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 47, 48, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 23, 24, 25, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, …, 58}
	KL-div (bits)	0.035	0.024	0.017	0.0079	0.002	0.0018	0.000096	0
	Entropy (bits)	4.94	5.0	4.97	5.0	5.07	5.05	5.13	5.13
	$(p, \hat{p})$	$(0.181, 0.070)$	$(0.170, 0.072)$	$(0.167, 0.075)$	$(0.157, 0.075)$	$(0.146, 0.072)$	$(0.145, 0.073)$	$(0.138, 0.070)$	$(0.137, 0.072)$
Thursday June 8, 2006	Total # of selected links	18	26	37	44	51	56	56	58
	$\mathcal{C}_{\text{IP}}$	{3, 4, 7, 8, 13, 14, 21, 22, 23, 24, 30, 32, 37, 38, 41, 44, 52, 58}	{2, 3, 4, 7, 8, 10, 11, 13, 14, 15, 17, 22, 23, 24, 30, 32, 33, 37, 38, 39, 43, 44, 52, 56, 57, 58}	{2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 22, 23, 24, 25, 28, 30, 32, 33, 35, 36, 37, 38, 41, 43, 44, 45, 46, 48, 51, 52, 53, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 22, 23, 24, 26, 27, 29, 30, 31, 32, 33, 35, 36, 37, 38, 40, 42, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 23, 24, 25, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, …, 58}
	KL-div (bits)	0.044	0.019	0.0092	0.008	0.0019	0.0003	0.00026	0
	Entropy (bits)	5.16	5.20	5.17	5.18	5.19	5.25	5.25	5.25
	$(p, \hat{p})$	$(0.182, 0.061)$	$(0.161, 0.063)$	$(0.150, 0.065)$	$(0.150, 0.065)$	$(0.145, 0.066)$	$(0.138, 0.064)$	$(0.138, 0.064)$	$(0.136, 0.064)$

Table 3.6: Solution of the set covering problem, entropy, KL-divergence and parameter of geometric distribution for Friday and Saturday and different δ^f values.

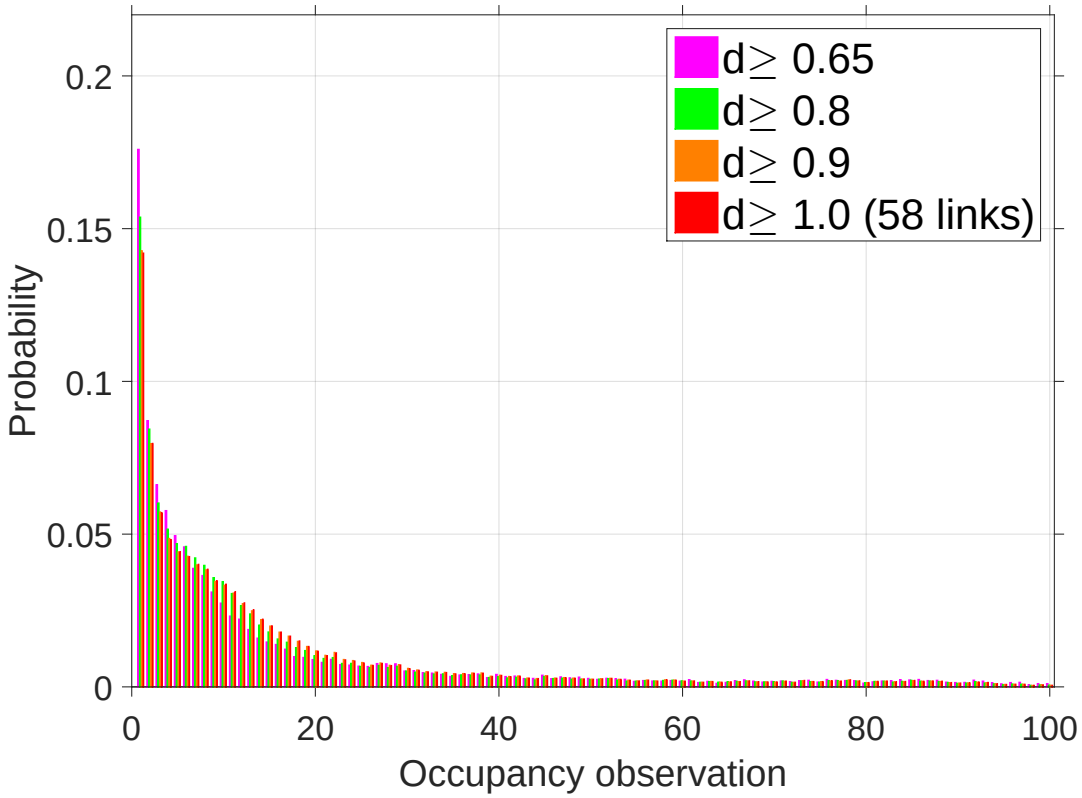
Day		"distance" ($d^f \geq \delta^f$)							
		$d^f \geq 0.6$	$d^f \geq 0.65$	$d^f \geq 0.7$	$d^f \geq 0.75$	$d^f \geq 0.8$	$d^f \geq 0.85$	$d^f \geq 0.9$	$d^f \geq 1.0$
Friday June 9, 2006	Total # of selected links	20	27	36	43	50	55	56	58
	$\mathcal{C}_{\text{IP}}$	{2, 4, 8, 11, 14, 16, 17, 22, 23, 24, 25, 30, 32, 37, 38, 41, 44, 52, 57, 58}	{2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 19, 21, 23, 24, 29, 30, 32, 33, 37, 38, 42, 44, 49, 52, 56, 57, 58}	{1, 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 18, 19, 20, 22, 23, 24, 25, 29, 30, 32, 33, 37, 38, 40, 41, 42, 44, 45, 51, 52, 53, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 22, 23, 24, 26, 27, 29, 30, 32, 33, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 51, 52, 53, 54, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 18, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, …, 58}
	KL-div (bits)	0.022	0.012	0.0078	0.0059	0.003	0.0006	0.00027	0
	Entropy (bits)	5.49	5.29	5.35	5.25	5.25	5.3	5.34	5.34
	$(p, \hat{p})$	$(0.134, 0.052)$	$(0.148, 0.059)$	$(0.136, 0.057)$	$(0.144, 0.062)$	$(0.141, 0.063)$	$(0.133, 0.061)$	$(0.131, 0.059)$	$(0.129, 0.060)$
Saturday June 10, 2006	Total # of selected links	31	38	44	48	52	56	57	57
	$\mathcal{C}_{\text{IP}}$	{1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 15, 17, 23, 24, 30, 31, 32, 33, 37, 38, 40, 42, 44, 48, 50, 52, 54, 56, 57, 58}	{1, 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 19, 23, 24, 27, 28, 30, 32, 33, 35, 37, 38, 40, 42, 44, 45, 46, 47, 48, 51, 52, 53, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 21, 22, 23, 24, 26, 27, 29, 30, 32, 33, 35, 36, 37, 38, 39, 40, 42, 44, 45, 46, 47, 48, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 30, 32, 33, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 18, 19, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, …, 49, 51, 52, …, 57}
	KL-div (bits)	0.024	0.03	0.0082	0.0039	0.0028	0.0003	0.00013	0.00013
	Entropy (bits)	4.88	4.76	4.87	4.89	4.9	4.98	4.99	4.99
	$(p, \hat{p})$	$(0.145, 0.079)$	$(0.156, 0.086)$	$(0.140, 0.082)$	$(0.140, 0.082)$	$(0.136, 0.082)$	$(0.129, 0.078)$	$(0.127, 0.078)$	$(0.126, 0.078)$

Table 3.7: Solution of the set covering problem, entropy, KL-divergence and parameter of geometric distribution for Sunday and different δ^f values.

Day		"distance" ($d^f \geq \delta^f$)							
		$d^f \geq 0.6$	$d^f \geq 0.65$	$d^f \geq 0.7$	$d^f \geq 0.75$	$d^f \geq 0.8$	$d^f \geq 0.85$	$d^f \geq 0.9$	$d^f \geq 1.0$
Sunday June 11, 2006	Total # of selected links	35	39	42	46	50	55	57	58
	$\mathcal{C}_{\text{IP}}$	{1, 2, 3, 4, 7, 8, 11, 12, 13, 14, 17, 22, 23, 24, 27, 28, 30, 32, 33, 35, 36, 37, 42, 43, 44, 47, 48, 49, 50, 52, 53, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 11, 12, 13, 14, 17, 19, 22, 23, 24, 27, 29, 30, 32, 33, 34, 37, 38, 41, 42, 43, 44, 46, 47, 48, 49, 51, 52, 53, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 14, 15, 18, 19, 22, 23, 24, 27, 29, 30, 32, 33, 34, 37, 38, 40, 41, 42, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 18, 22, 23, 24, 26, 27, 29, 30, 31, 32, 33, 35, 37, 38, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 21, 22, 23, 24, 25, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58}	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58}	{1, 2, …, 58}
	KL-div (bits)	0.043	0.031	0.023	0.014	0.0062	0.0005	0 .00014	0
	Entropy (bits)	4.15	4.19	4.27	4.27	4.3	4.39	4.43	4.43
	$(p, \hat{p})$	$(0.209, 0.128)$	$(0.198, 0.126)$	$(0.177, 0.120)$	$(0.177, 0.122)$	$(0.165, 0.121)$	$(0.151, 0.116)$	$(0.148, 0.113)$	$(0.147, 0.113)$

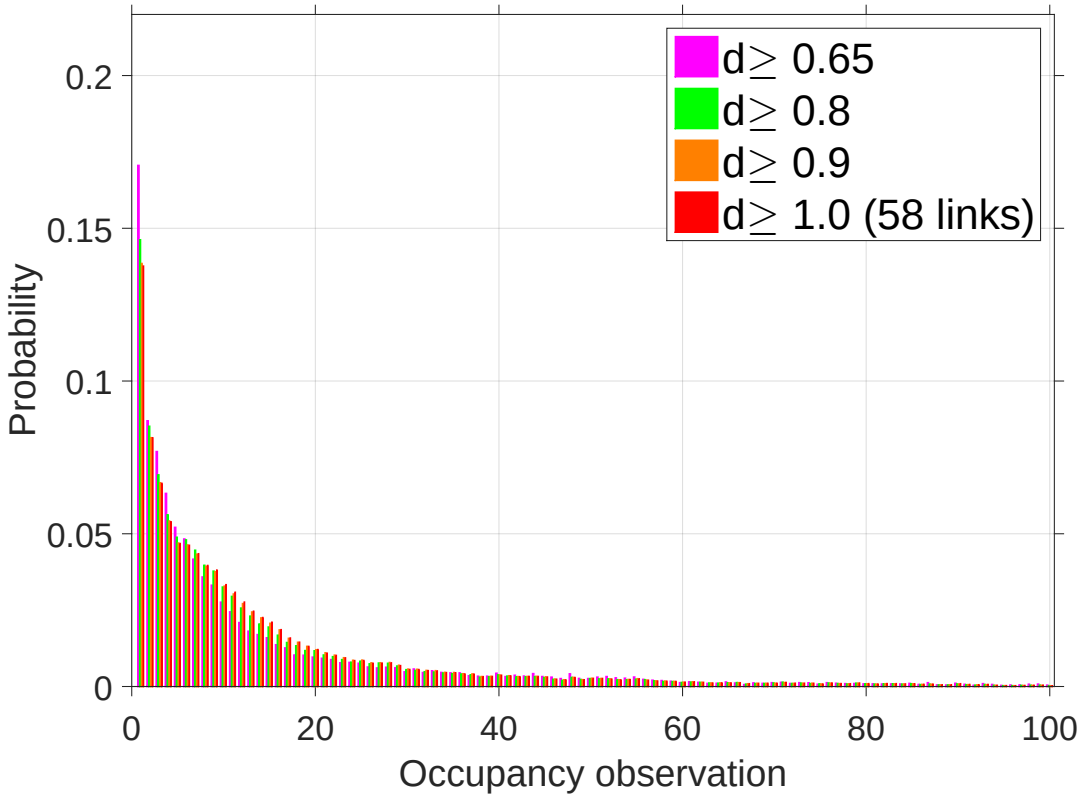


(a) Monday

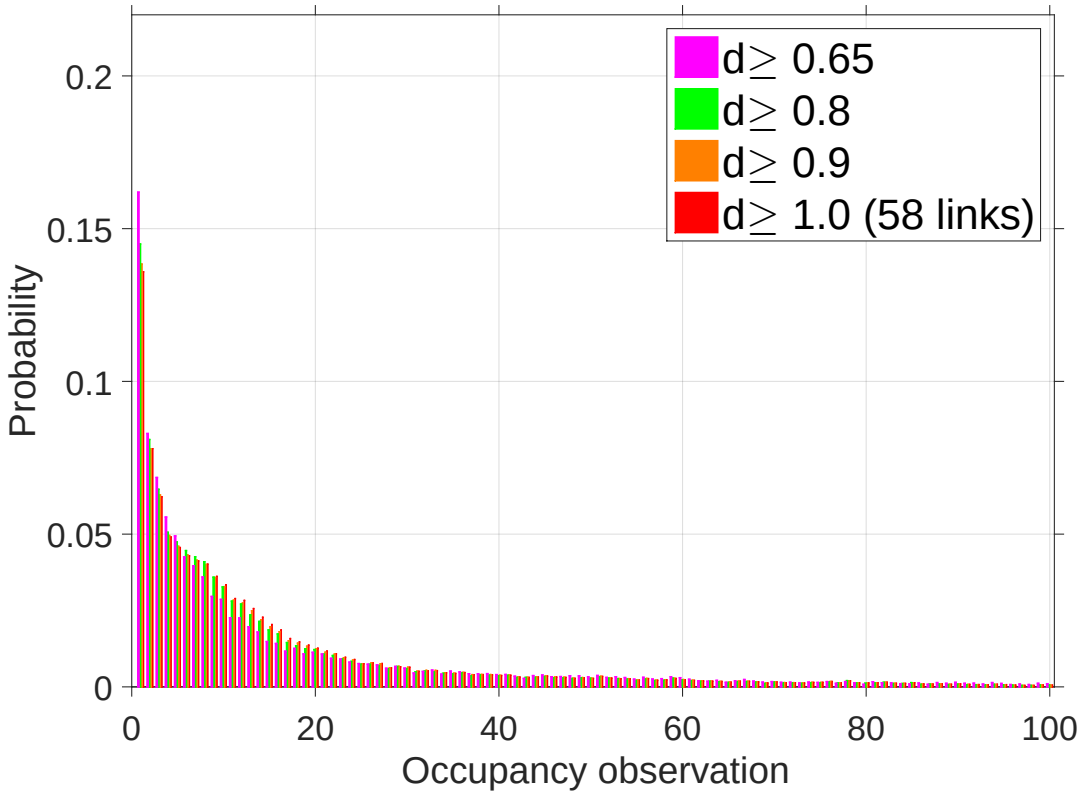


(b) Tuesday

Figure 3.11: Comparison between pmf and geometric distribution Monday and Tuesday.

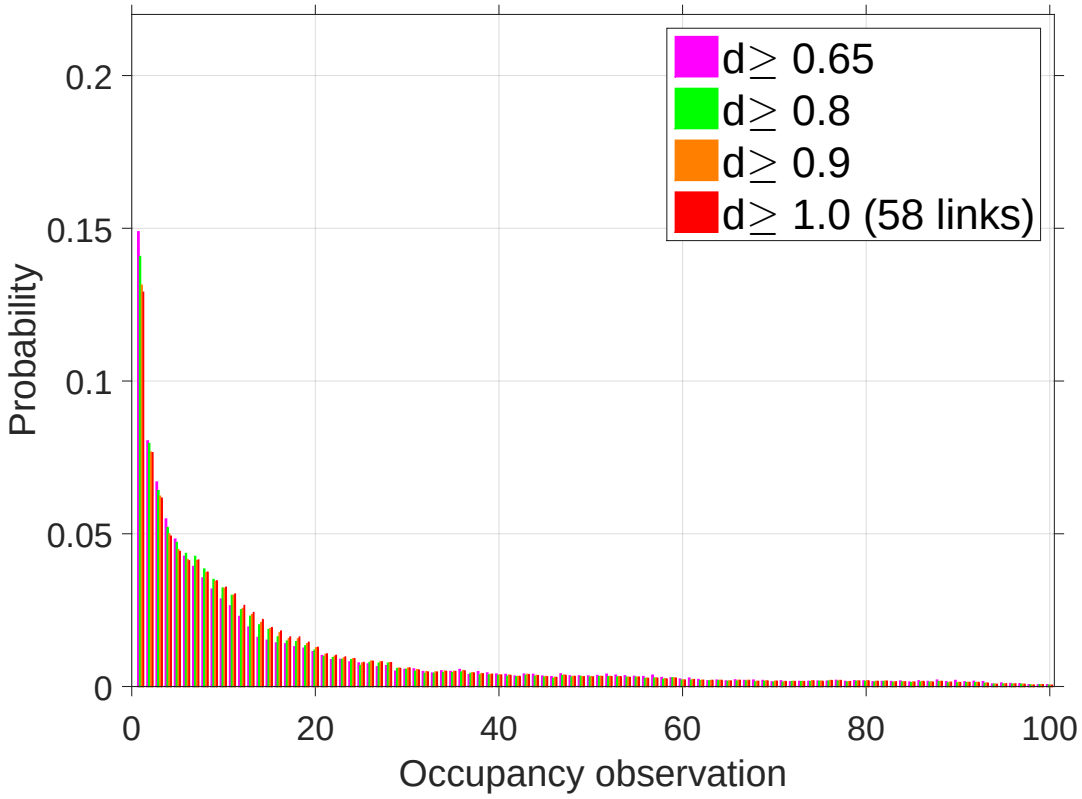


(a) Wednesday

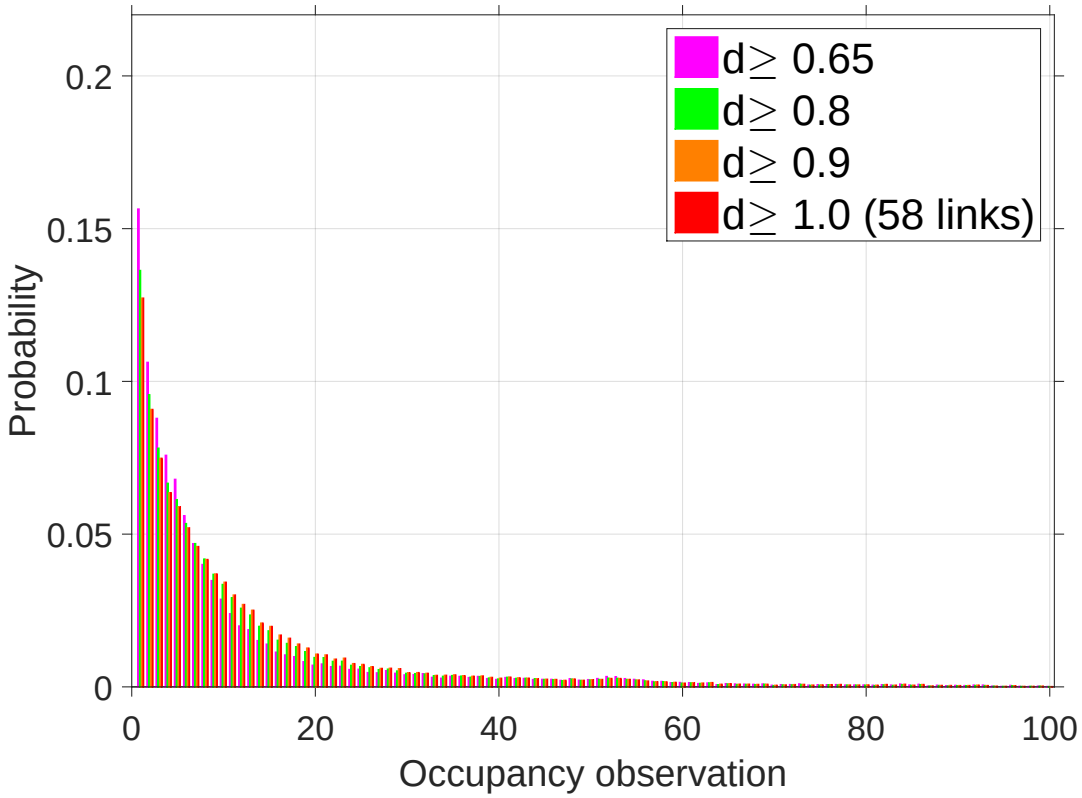


(b) Thursday

Figure 3.12: Comparison between pmf and geometric distribution Wednesday and Thursday.



(a) Friday



(b) Saturday

Figure 3.13: Comparison between pmf and geometric distribution Friday and Saturday.

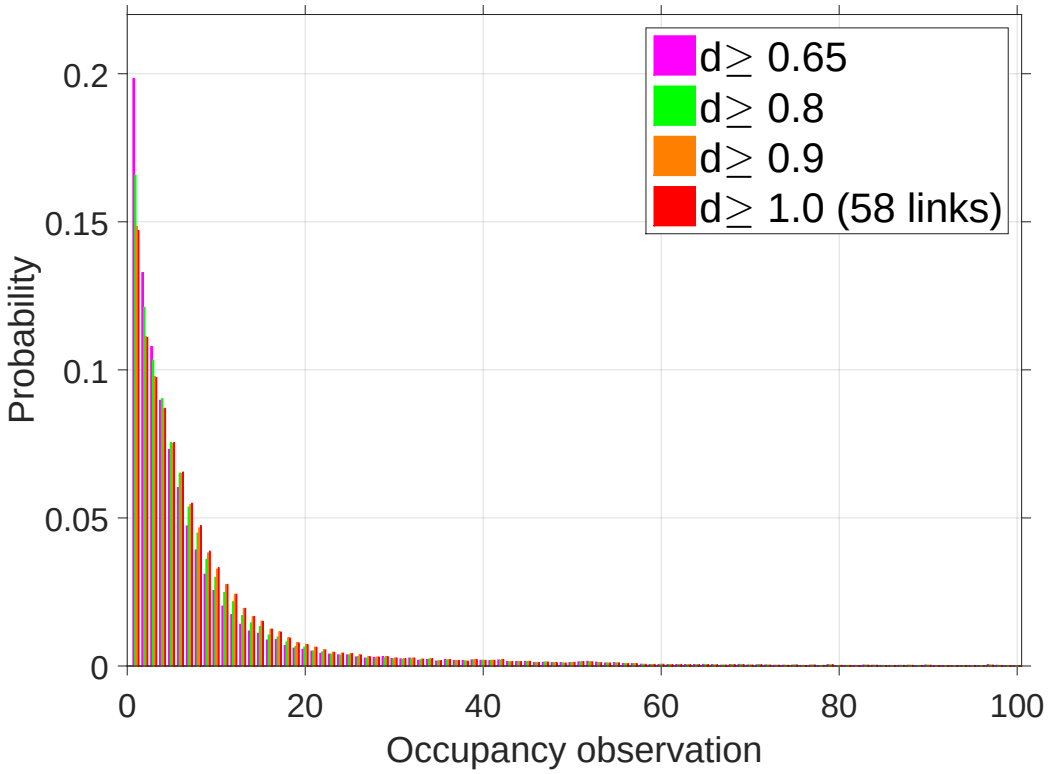


Figure 3.14: Comparison between pmf and geometric distribution Sunday.

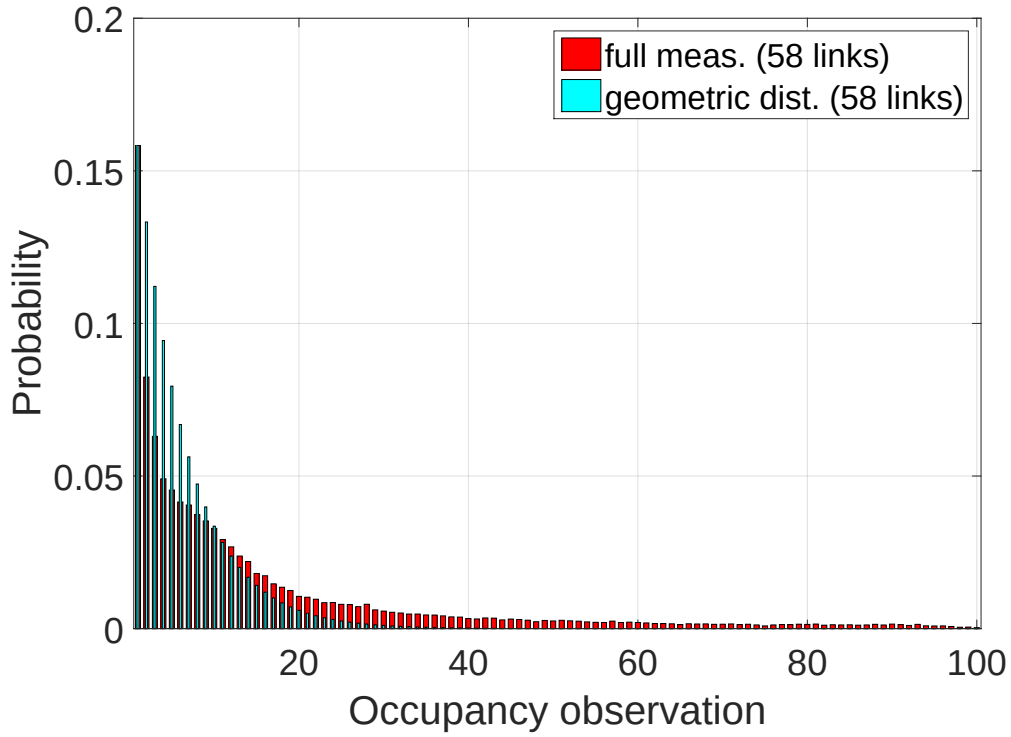
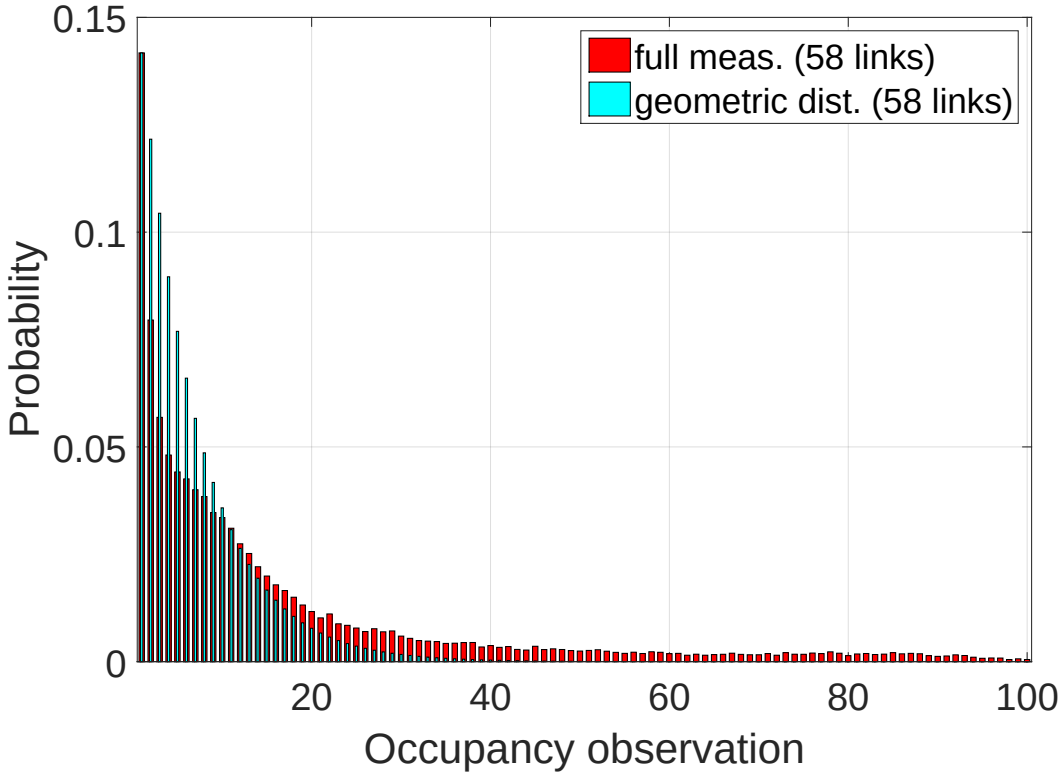
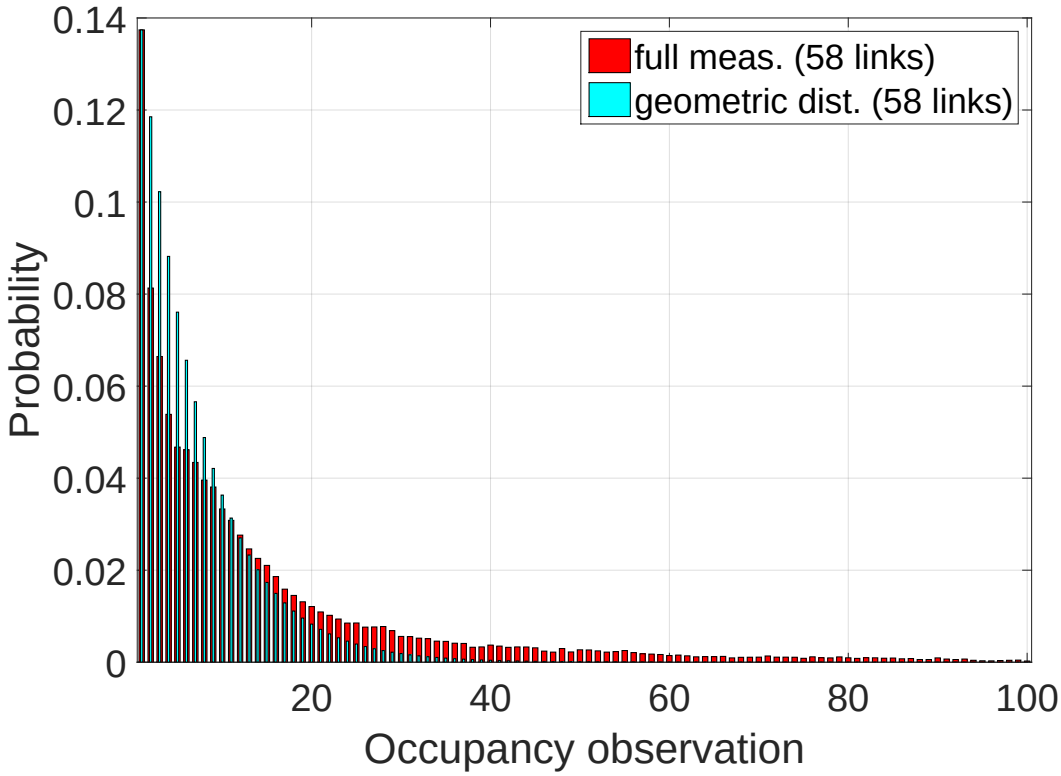


Figure 3.15: Comparison between pmf and geometric distribution at $d^f \geq 1.0$ Monday.

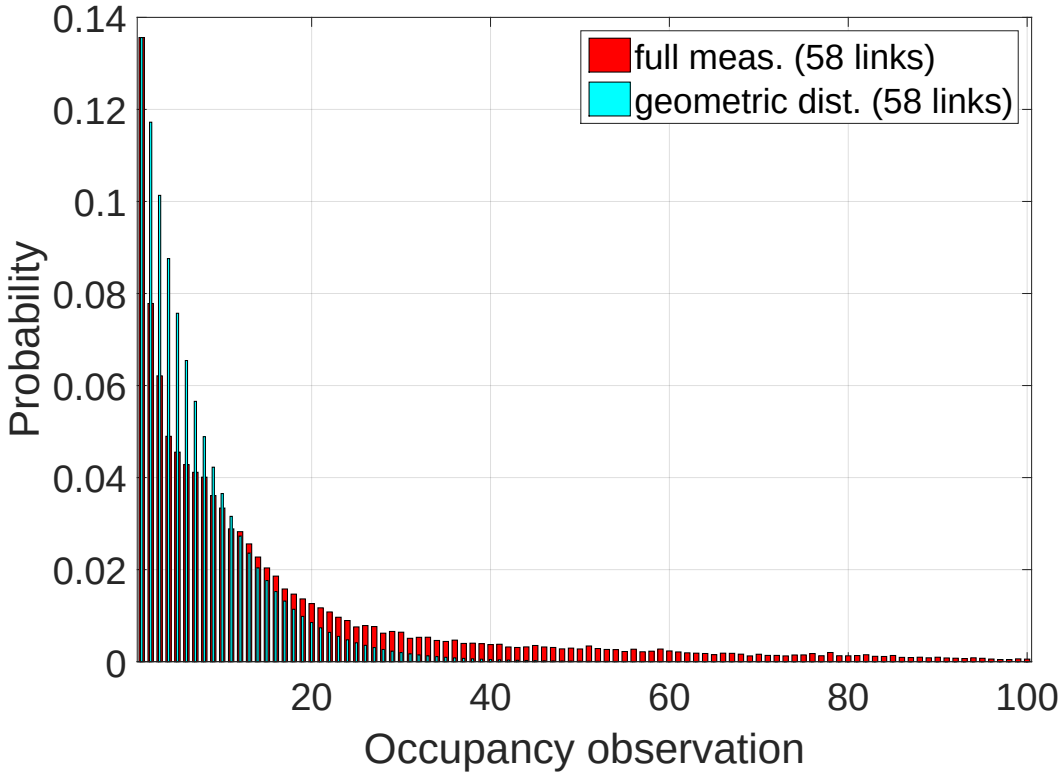


(a) Tuesday

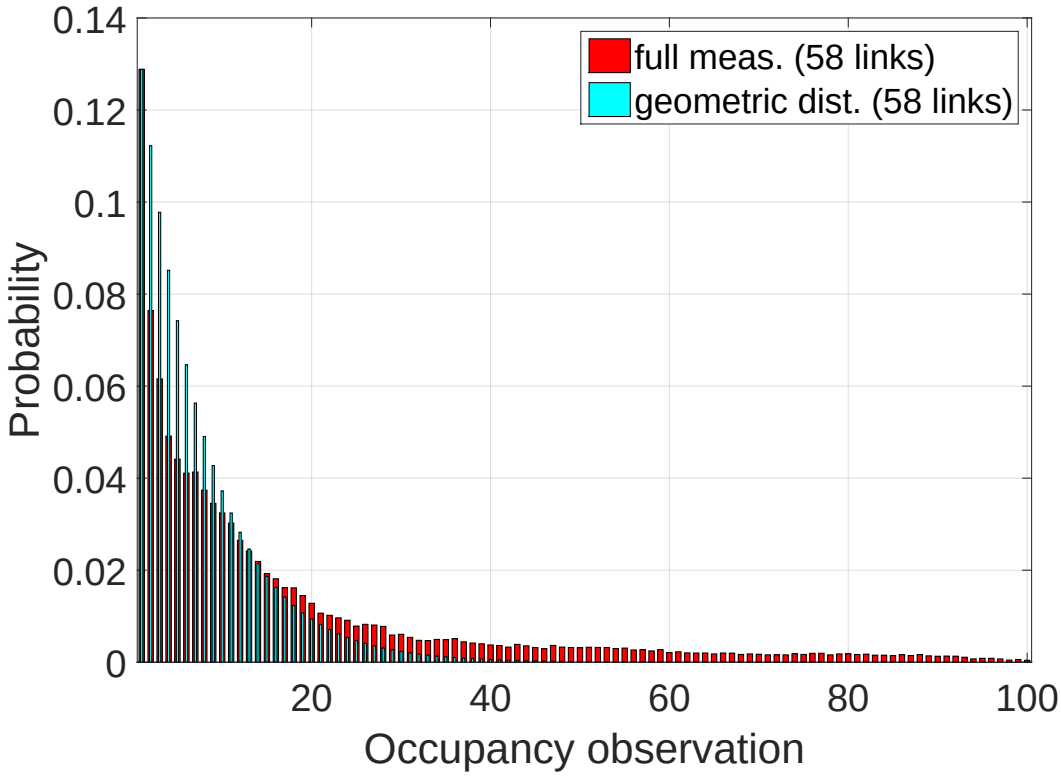


(b) Wednesday

Figure 3.16: Comparison between pmf and geometric distribution $d^f \geq 1.0$ Tuesday and Wednesday.

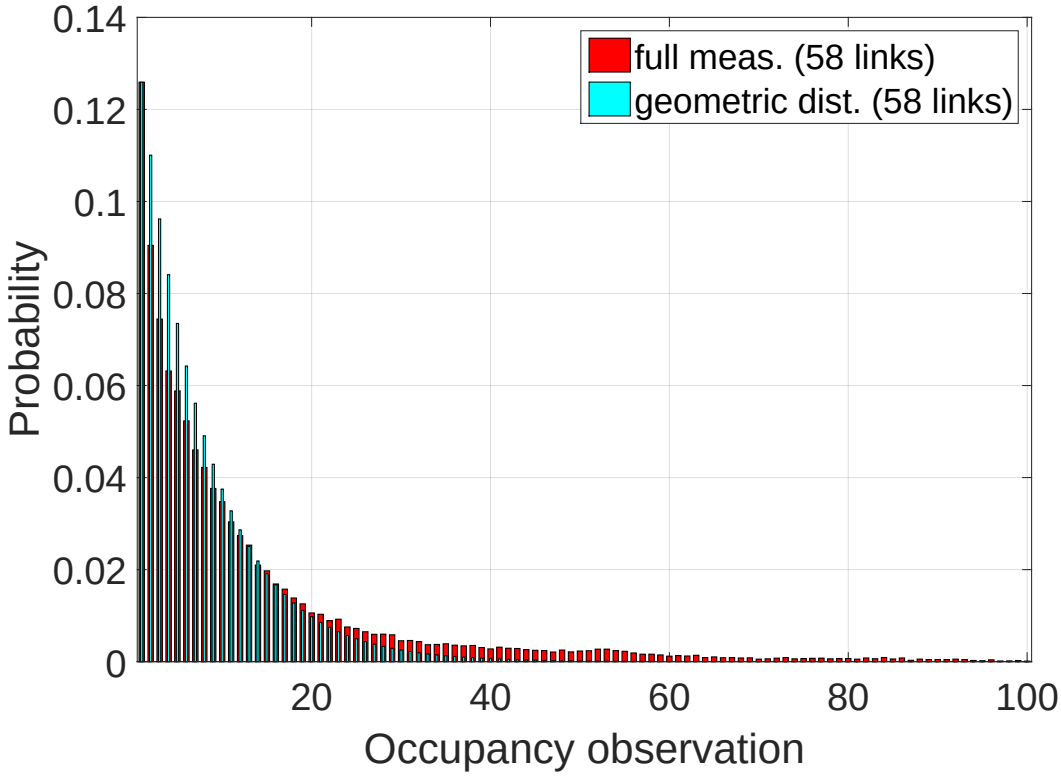


(a) Thursday

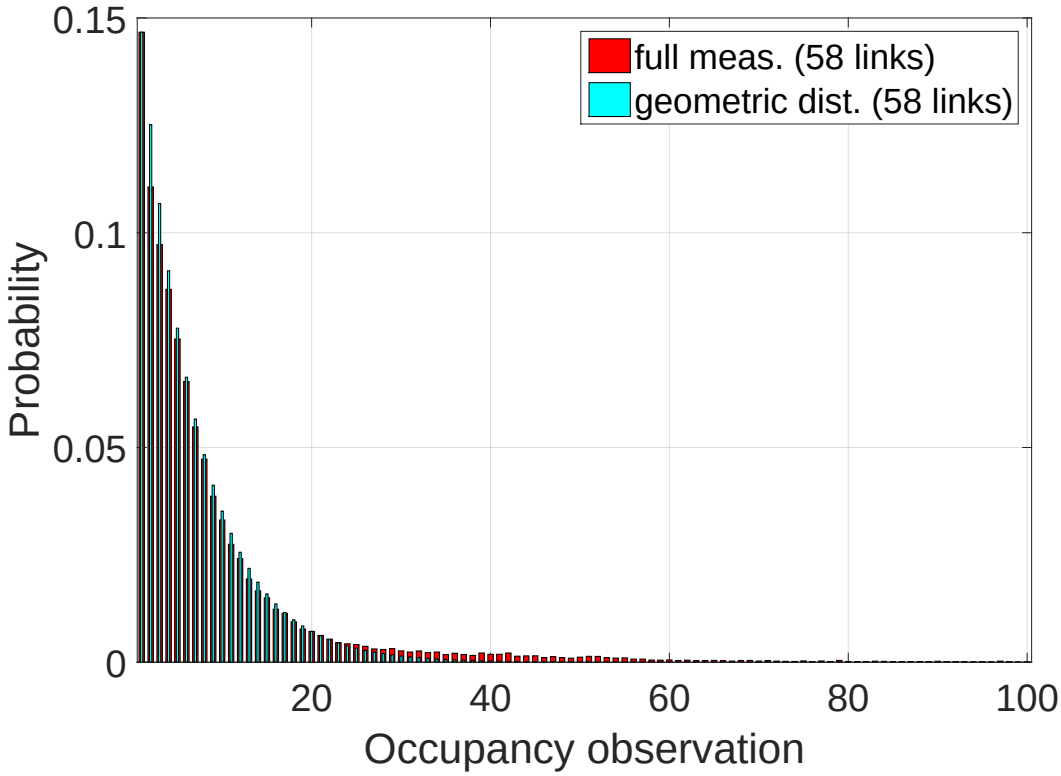


(b) Friday

Figure 3.17: Comparison between pmf and geometric distribution $d^f \geq 1.0$ Thursday and Friday.

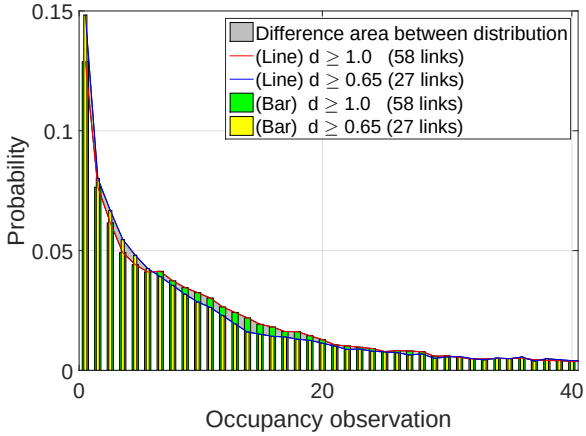


(a) Saturday

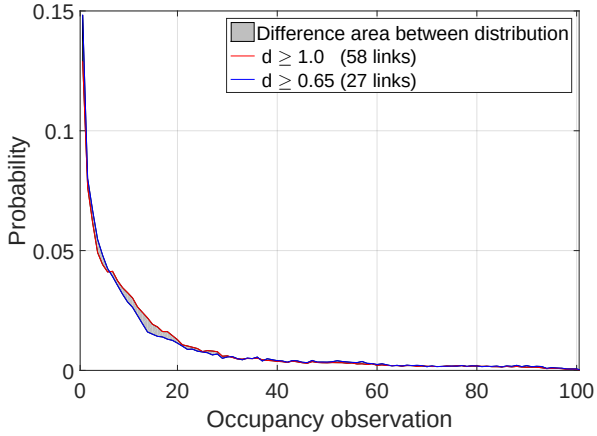


(b) Sunday

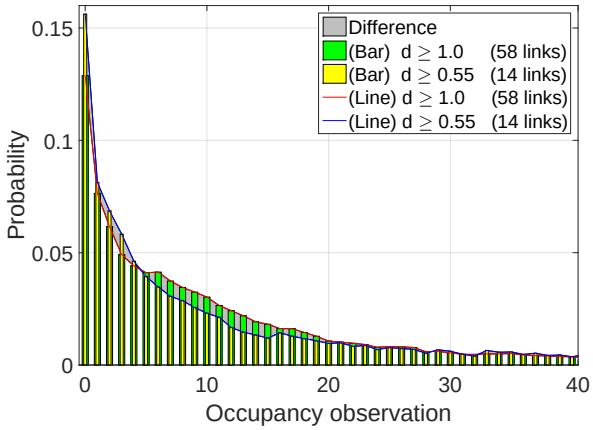
Figure 3.18: Comparison between pmf and geometric distribution $d^f \geq 1.0$ Saturday and Sunday.



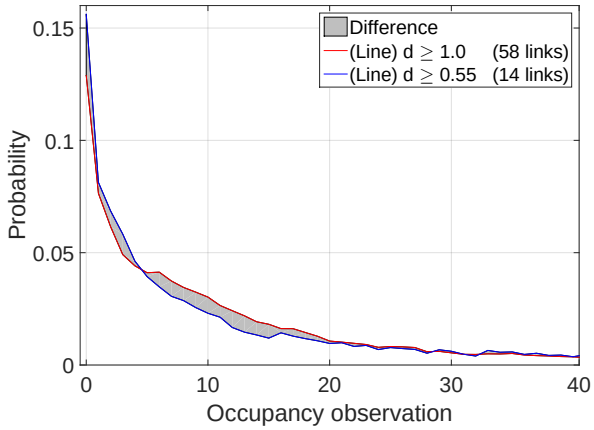
(a) Friday pmf at $d^f \geq 0.65$



(b) Line graph at $d^f \geq 0.65$

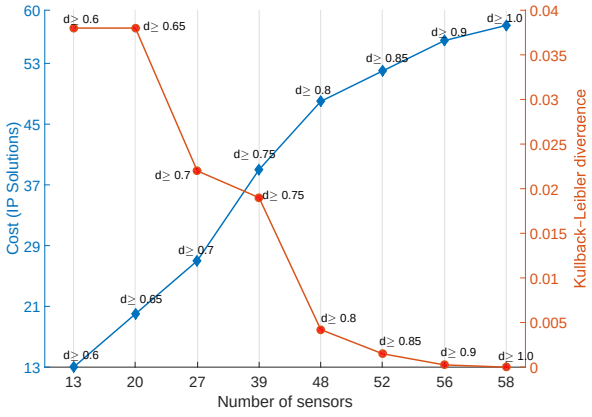


(c) Friday pmf at $d^f \geq 0.55$

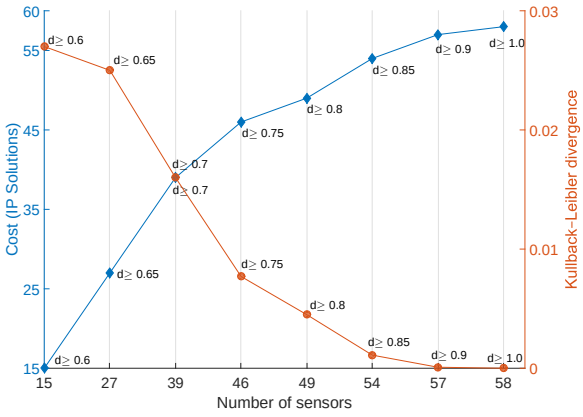


(d) Line graph at $d^f \geq 0.55$

Figure 3.19: Comparison Friday pmf.

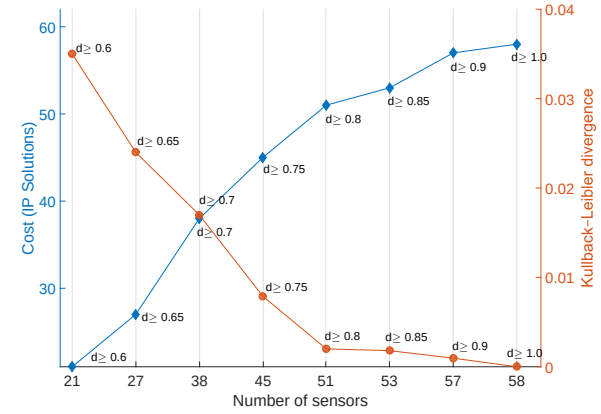


(a) Monday

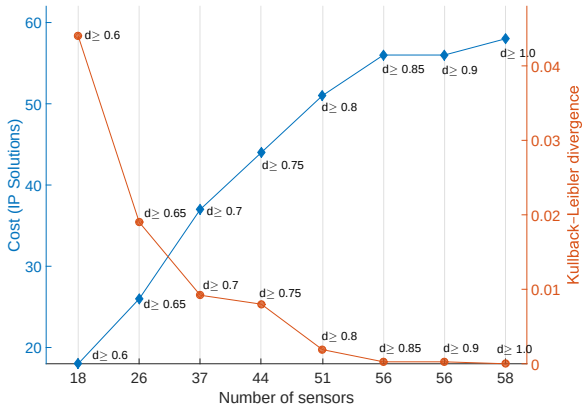


(b) Tuesday

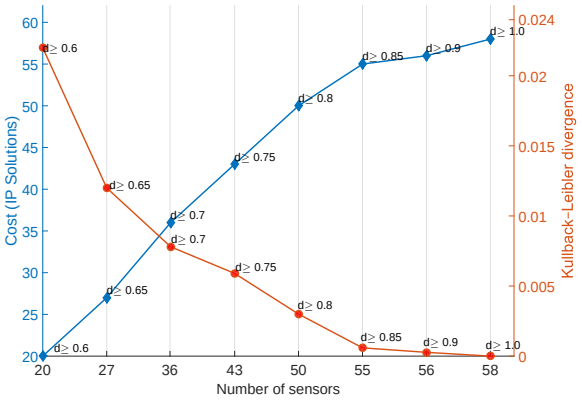
Figure 3.20: Kullback-Leibler divergence versus cost of the GIP Monday and Tuesday.



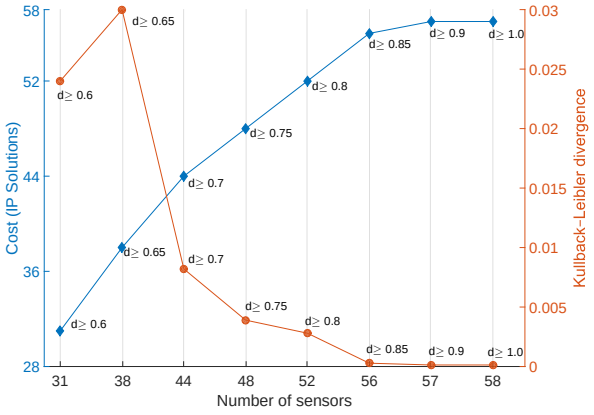
(a) Wednesday



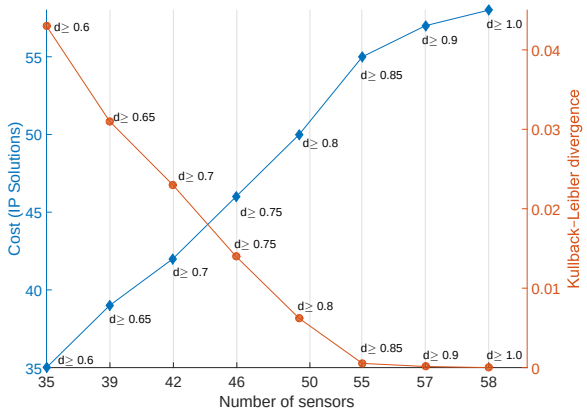
(b) Thursday



(c) Friday

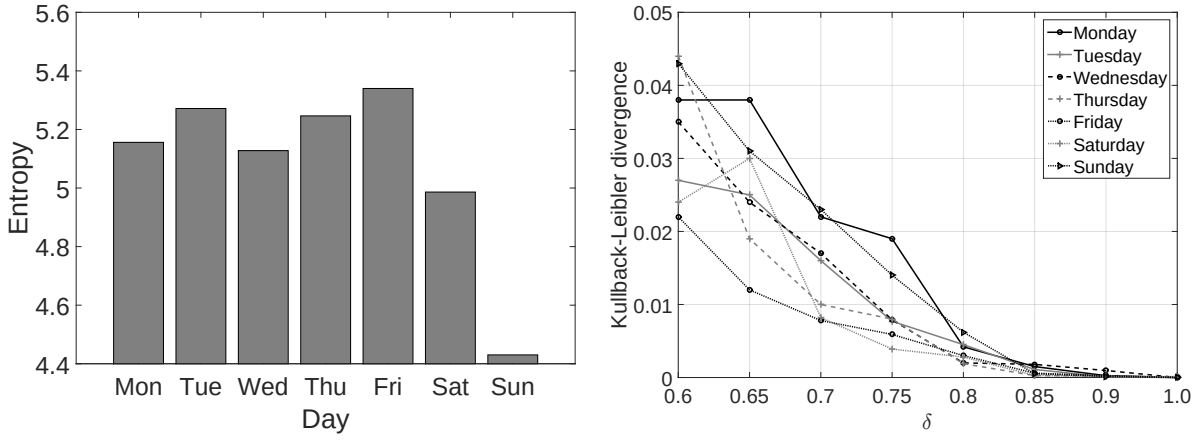


(d) Saturday



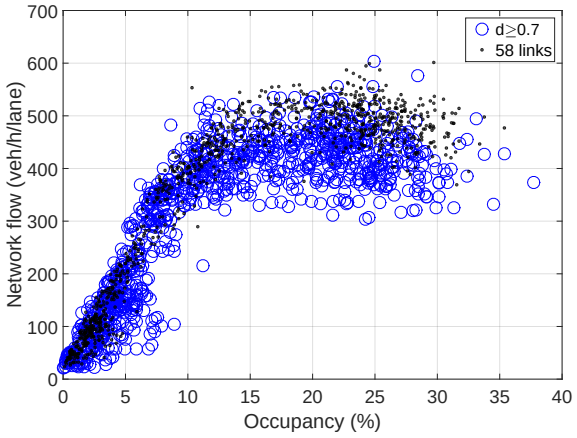
(e) Sunday

Figure 3.21: Kullback-Leibler divergence versus cost of the GIP Wednesday to Sunday.

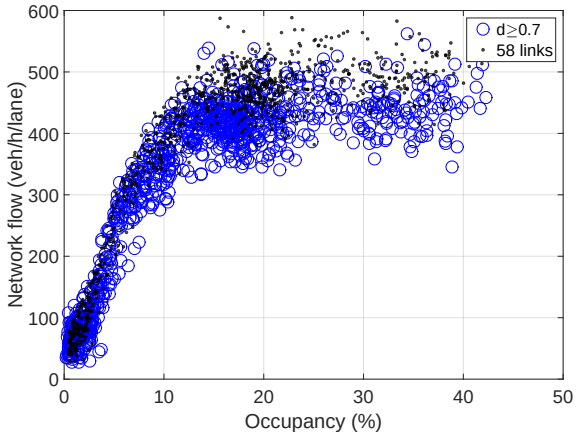


(a) (b)

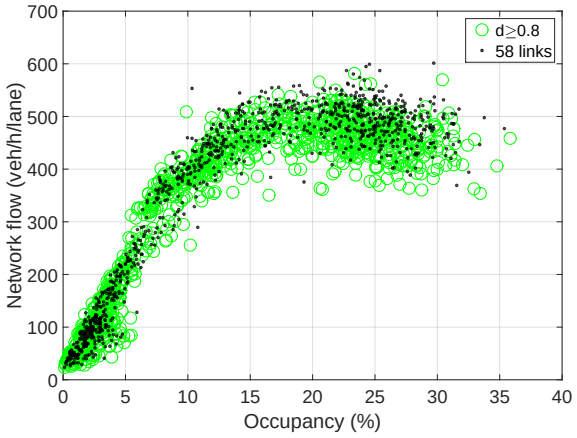
Figure 3.22: (a) Information entropy (in bits) of each day of a week; (b) Kullback-Leibler divergence in function of δ for one week.



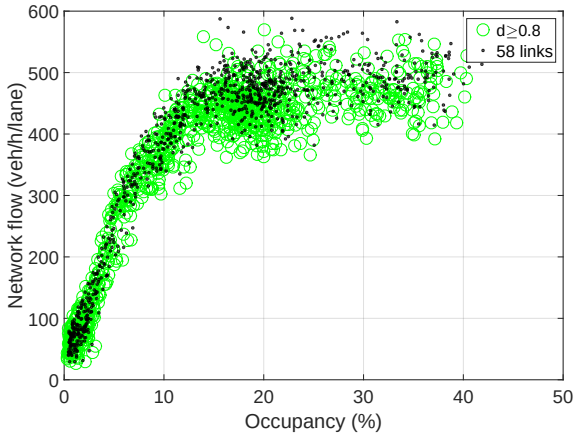
(a) Monday $d \geq 0.7$ (27 links)



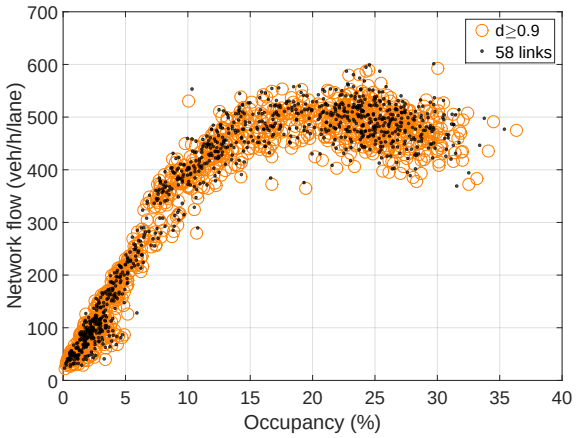
(d) Tuesday $d \geq 0.7$ (39 links)



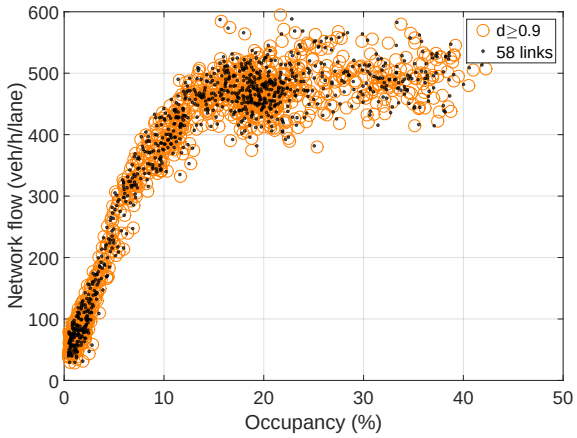
(b): Monday $d \geq 0.8$ (48 links)



(e): Tuesday $d \geq 0.8$ (49 links)

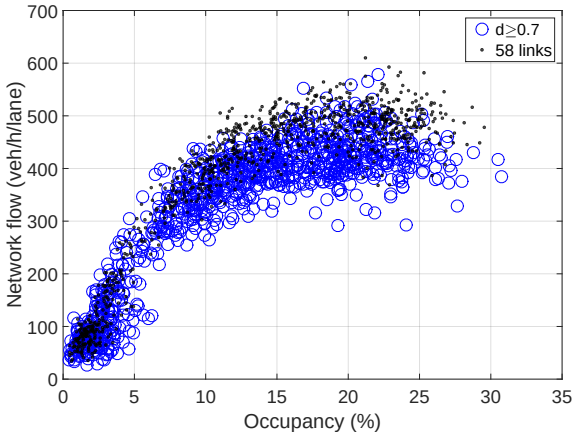


(c): Monday $d \geq 0.9$ (56 links)

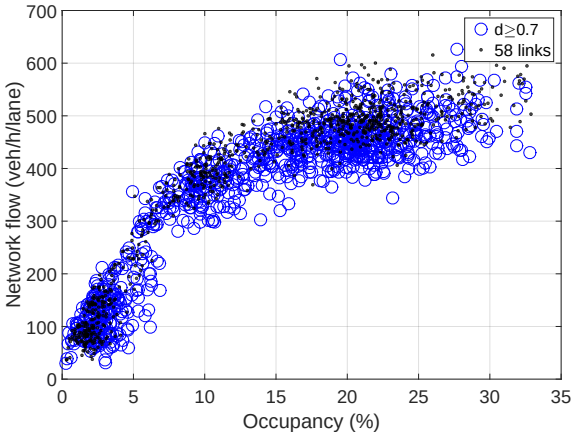


(f): Tuesday $d \geq 0.9$ (57 links)

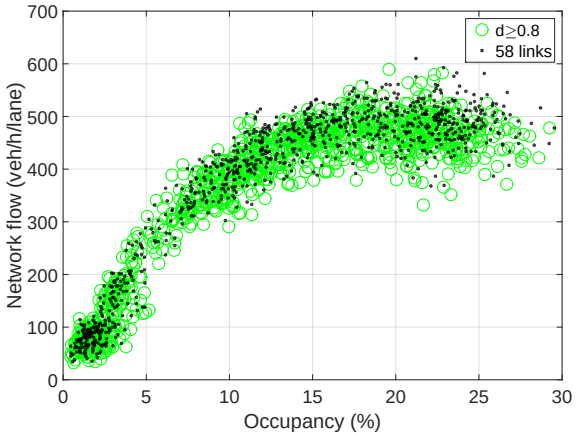
Figure 3.23: Sparse-measurement NFD approximations Monday and Tuesday and different δ values.



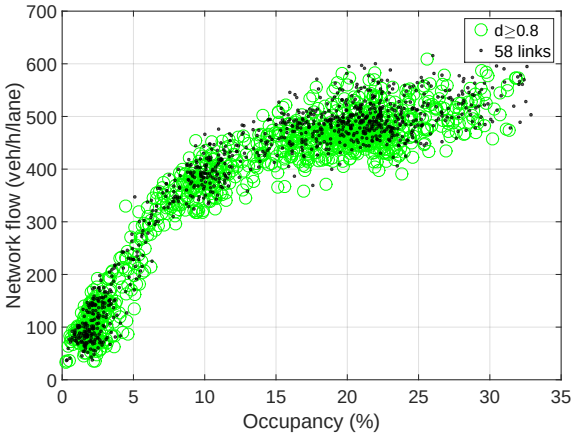
(a) Wednesday $d \geq 0.7$ (38 links)



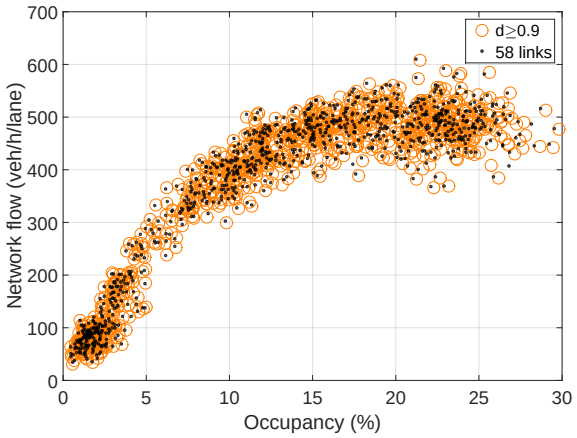
(d) Thursday $d \geq 0.7$ (37 links)



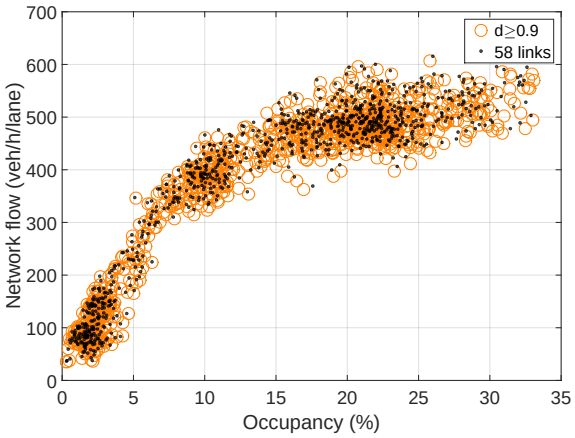
(b): Wednesday $d \geq 0.8$ (51 links)



(e): Thursday $d \geq 0.8$ (51 links)

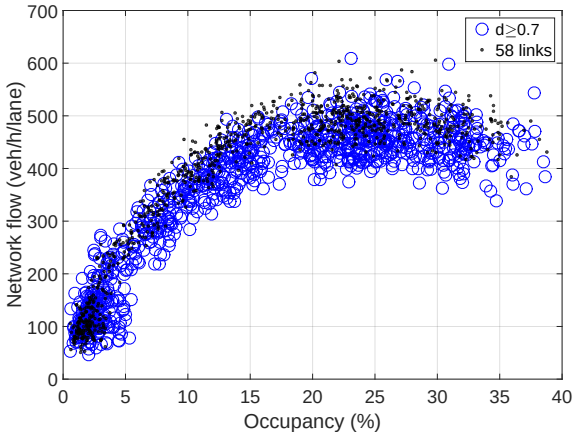


(c): Wednesday $d \geq 0.9$ (57 links)

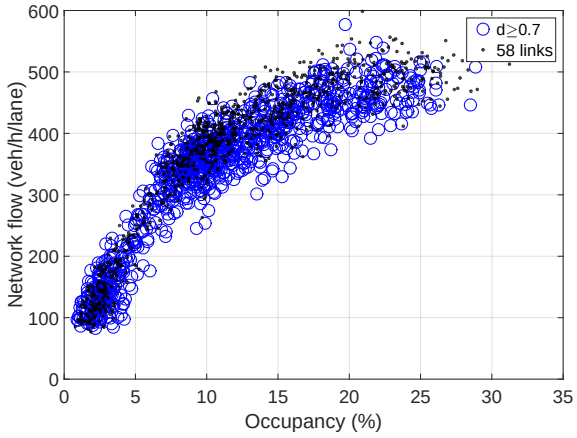


(f): Thursday $d \geq 0.9$ (56 links)

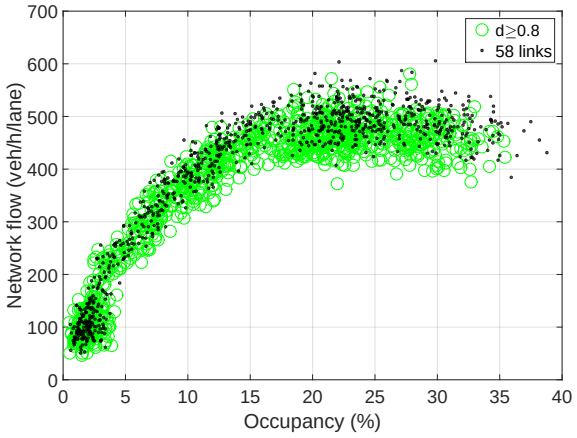
Figure 3.24: Sparse-measurement NFD approximations Wednesday and Thursday and different δ values.



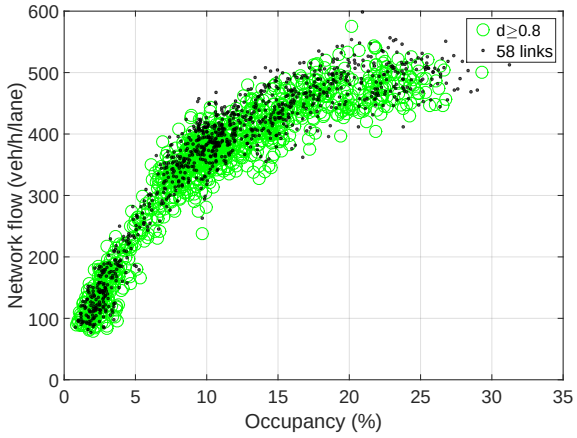
(a) Friday $d \geq 0.7$ (36 links)



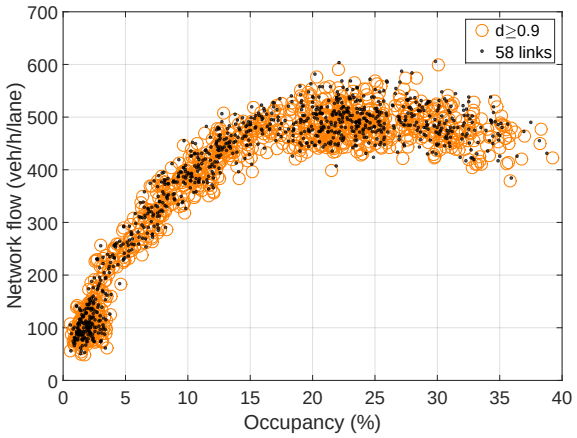
(d) Saturday $d \geq 0.7$ (44 links)



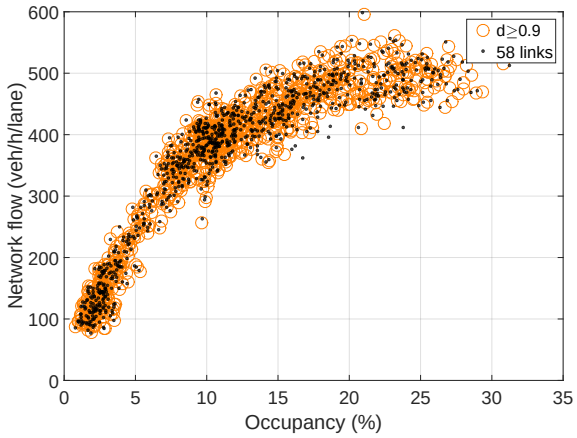
(b): Friday $d \geq 0.8$ (50 links)



(e): Saturday $d \geq 0.8$ (52 links)

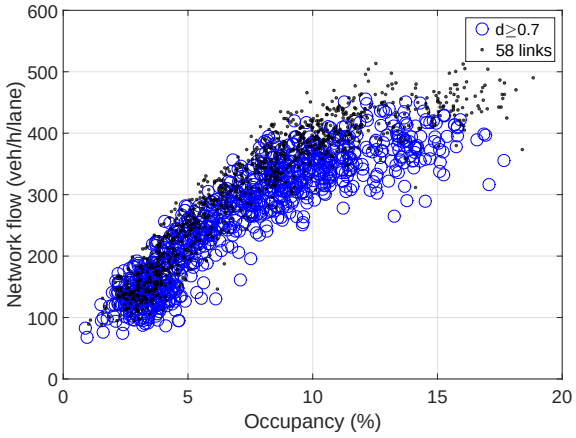


(c): Friday $d \geq 0.9$ (56 links)

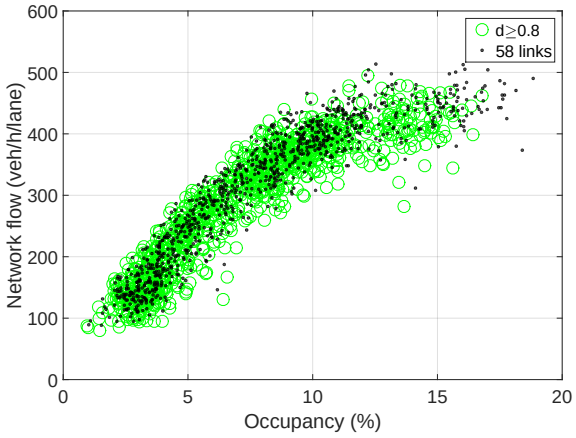


(f): Saturday $d \geq 0.9$ (57 links)

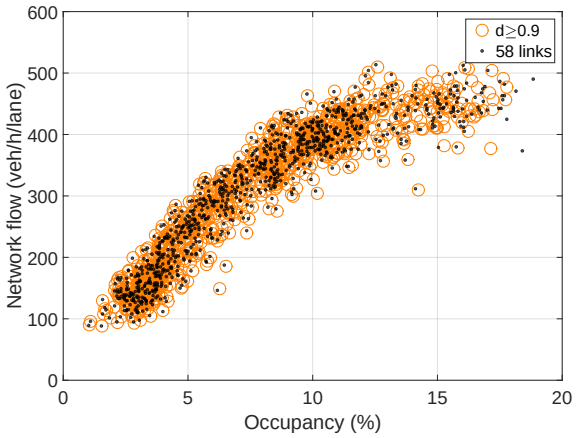
Figure 3.25: Sparse-measurement NFD approximations Friday and Saturday and different δ values.



(a) Sunday $d \geq 0.7$ (42 links)



(b): Sunday $d \geq 0.8$ (50 links)



(c): Sunday $d \geq 0.9$ (57 links)

Figure 3.26: Sparse-measurement NFD approximations Sunday and different δ values.

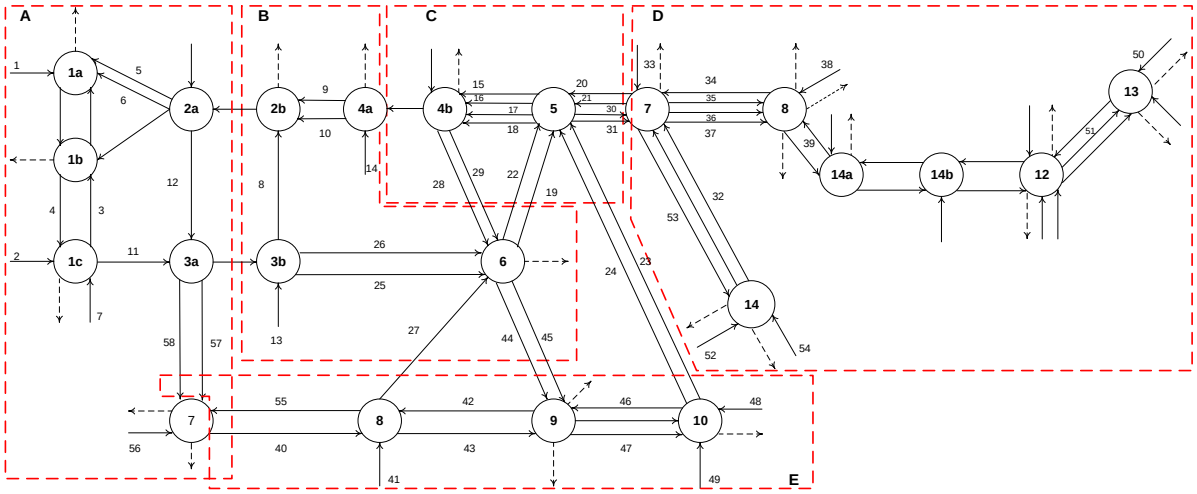


Figure 3.27: Partition of the network into five regions (A, B, C, D, and E).

3.3.4 Ad-hoc strategies

To test the effectiveness of the GIP formulation, two alternative strategies for sensor selection are considered. The first strategy is called "Blind strategy" and sensors selected randomly. The second strategy is called "Radius strategy" and the spatial distance is used as a metric. The main features of the two strategies are as follows:

Blind strategy: Blind selection of sensors is cheap and usually provides good level of approximation. More importantly, the sensor selection problem is in general a stochastic optimisation problem involving noisy measurements. In principle, there is countably infinite number of solutions (sensor combinations) that can provide similar performance. In our context, links are selected randomly and the NFD constructed from their flow and occupancy measurements.

Radius strategy: The radius strategy does not genuinely reflects the spatial distance. Mainly it selects sensors based on the topology of the network. The idea here is to identify if sensors located at specific regions of the network contribute to the better approximation of the NFD. For example in Figure 3.27, the network is partitioned into five regions (A, B, C, D, and E) consisting of the same number of sensors for fair comparisons.

3.3.5 Comparative study

In the previous sections, it is demonstrated that the basic GIP model and the proposed information-based methodology is able to determine the set of links/sensors to construct an operational or sparse-measurement NFD. To further demonstrate the effectiveness of the proposed framework, the coverage level of the network is reduced to 24% (14 links) and a comparative study of a number of models obtained from the GIP model with blind and radius strategies is

Table 3.8: Comparison of different models involving less number of links for Friday.

Parameters	Full-measurement NFD (Reference NFD)	Model 2	Model 3	Model 4	Model 5 (Blind)	Model 6 (Radius C)	Model 7 (Radius E)
δ^o	-	0.37	-	0.15	-	-	-
δ^f	-	0.51	0.32	0.25	-	-	-
R	-	-	100	100	-	-	-
$\mathcal{C}$	58 links	{2,4,19,23, 24, 30,37,39, 41,44,48,52, 55,57}	{2,3,6,7, 12,14,18,23, 34,41,45,48, 51,57}	{2,4,5,7, 9,14,18,22, 23,42,47,50, 53,58}	{6,11,12,21, 24,28,32,34, 35,36,40,49, 54,55}	{15,16,17,18, 19,20,21,22, 23,24,28,29, 30,31}	{27,40,41,42, 43,44,45,46, 47,48,49,55, 57,58}
Crit. occupancy (%) +/- 5%	25 (20 - 30)	25	23	24	26	33	25
Flow capacity (veh/h) +/- 10%	520 (470 - 570)	515	508	557	580	574	508
RMSE	38	74	62	81	72	73	70
KL-div (bits)	0	0.045	0.014	0.007	0.014	0.039	0.031

performed. To this end, six different strategies were evaluated: three GIP-based models (Models 2, 3, and 4), blind strategy (Model 5), and two variation of the radius strategy (Model 6 and 7). The two variation of the radius strategy refer to the areas C and E in Figure 3.27). To simplify the representation of results, the analysis will hereafter be limited to the data set of Friday.

Table 3.8 summarises the obtained results for the six different models. Four criteria were used for assessing the performance of the different models, including two features of the full-measurement NFD as reference (critical occupancy range, flow capacity range) and two metrics (root mean square error (RMSE) and KL-divergence). δ^o and δ^f were selected by a trial-and-error procedure to achieve a GIP solution with 14 sensors/links, while R was set to 100 m. RMSE is calculated between the fitted reference NFD (full measurement NFD) and the observed values of the model. The acceptable error for critical occupancy and flow capacity are set to 5% and 10%, respectively. KL-divergence is calculated from (A.2).

As can be seen in Table 3.8, models 2, 3, 4, and 7 manage to preserve the characteristics of the reference NFD. Their critical occupancy and flow capacity are observed to be within the acceptable error with minimal or zero deviation. Model 5 and 6 indicate higher flow capacity than the acceptable range. Model 3 exhibits the lowest RMSE when contrasted with Model 2. Remarkably, Model 3 indicates small RMSE despite the absence of occupancy as parameter in GIP. Model 4 achieves the best KL-divergence value compared to all models, but with high RMSE. This can be attributed to the uncertainty involved if a small δ^o is used. Model 4 excellent performance is also attributed to the combination of all (three) sources of information in the GIP formulation. For the radius strategy, the two selected areas indicated contradicting results. As can be seen in the table, Model 6 is far from the reference NFD. In contrast, Model 7 shows good agreement with the reference NFD as well as good RMSE and KL-div. Note that Model 6 represents the traffic dynamics in city centre, while Model 7 represents an arterial. Concluding, the proposed GIP models guarantee information accuracy for different levels of network coverage.

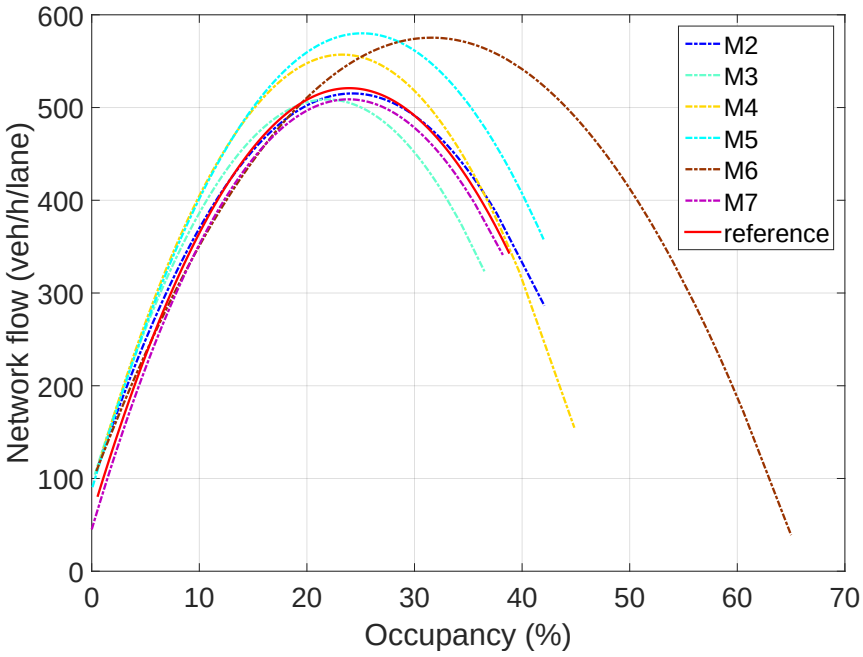


Figure 3.28: Fit line of different models.

Figure 3.30 shows the sparse-measurement NFDs for the six different models. As a first observation, a number of models using different sources of data can be used to approximate the full-measurement NFD. By visual inspection, besides Figure 3.30(e), most of the models provide good approximation as their shapes and scatter are in good agreement with the full measurement NFD, despite using less number of links. The proposed GIP models (Models 2, 3, and 4) manage to well-reproduce the congested regime, which is not the case in blind strategies (cf. Figure 3.30(c) with Figure 3.30(d)). The small RMSEs observed in Models 3 and 7 are explained in Figures 3.30(b) and 3.30(f), respectively. On the other hand, a significantly higher scatter with a wide dispersion of points up to the maximum occupancy is observed in Figure 3.30(e) (centre of the city). This underlines that installing sensors in the most congested links of the network does not guarantee a good approximation of the NFD.

Figure 3.30 also shows the best fit approximation of individual models and their comparison with the reference NFD. By visual inspection, Models 2, 3, 4, 5, and 7 give a good approximation. However, Model 6 (radius strategy) indicates a poor diagram which is consistent with the high RMSE and KL-divergence reported in Table 3.8. In some diagrams the high scatter is obvious (see e.g., Figure 3.30(c)). This attributed to information loss in models using a small number of links. Certainly, visual inspection is subjective and for this reason the accuracy of different models are assessed by the KL-divergence and RMSE, as explained in Table 3.8.

Figure 3.28 summarises the obtained models by providing their best fit curve approximation (a second order polynomial for all models is used). Models 2, 3, 4, and 7 fit well with the reference NFD (red curve) and preserve the critical occupancy between 20% and 30%. Models

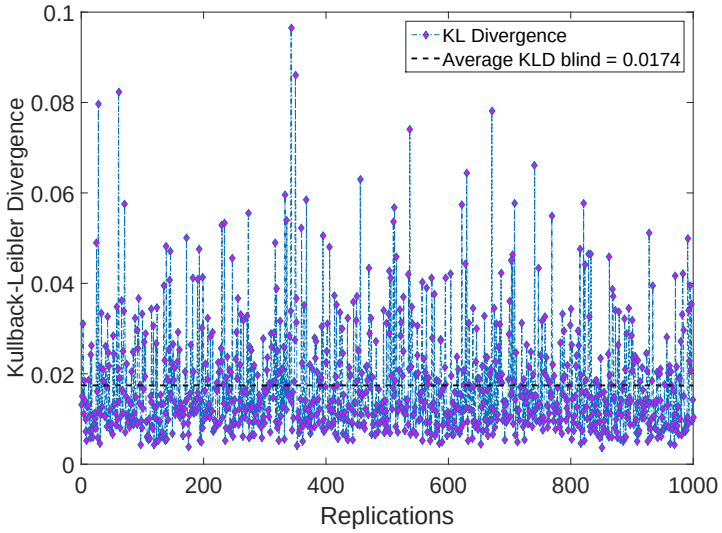
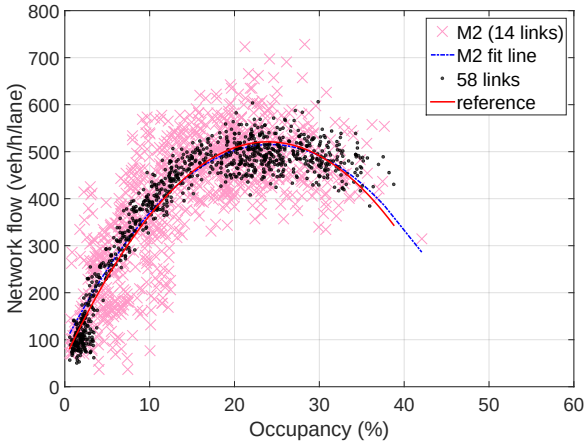


Figure 3.29: One thousand of replications involving twenty randomly selected links.

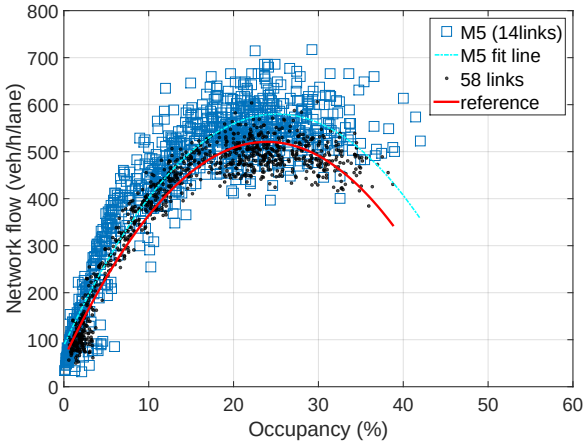
4 and 7 are better in representing the full measurements. This underlines that the proposed GIP strategy provides a good NFD approximation. On the other hand, curve 6 (representing the radius strategy - C area) indicates the worst fit. The blind strategy indicates a good fit though the capacity flow exceeds the acceptable error.

Figure 3.31 highlights the selected link or sensor selections of the comparative studies for Model 2, 3 and 4. As can be seen, locations of selected link (sensor) are vary between three models attributed to the different parameters used. It can be observed that without paramater radius R of 100m, Model 2 select links which are close to each other. For example in Figure 3.31(a), link 23 and 24 are nearby which could retain same coverage of traffic information. Figure 3.31(b)-(c) suggest that when the location of sensors are distributed 'uniformly' in the network, the information is greatly enhanced to account different traffic dynamics. The selected links are essential as they cover different areas of the network such as city centre, arterial, etc. As a result, Model 3 and 4 produced an improved performance of approximating fundamental diagram in comparison with Model 2.

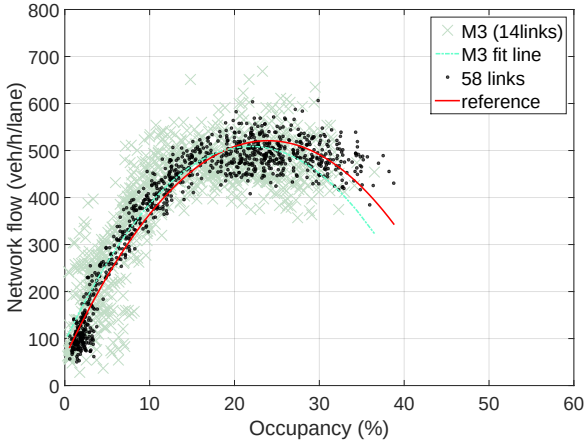
From the experiments, it is observed that a number of rival models can approximate the nominal NFD. To demonstrate this, 1000 replications is ran involving twenty (20) randomly selected links and their KL-divergence are calculated. Figure 3.29 displays the obtained results. As can be seen most of the replications indicate small KL-divergence values. In other words, many models can provide good information accuracy and thus efficiently approximate the NFD. This is attributed to the high stochasticity of the involved combinatorial optimisation problem. One other possible reason might be the small size of network used in this study. The availability of real empirical data for this network though give us some good evidence on the approximation problem. It is expected that a larger heterogeneous network with multiple centers of congestion would make the performance of blind or other ad-hoc strategies to deteriorate.



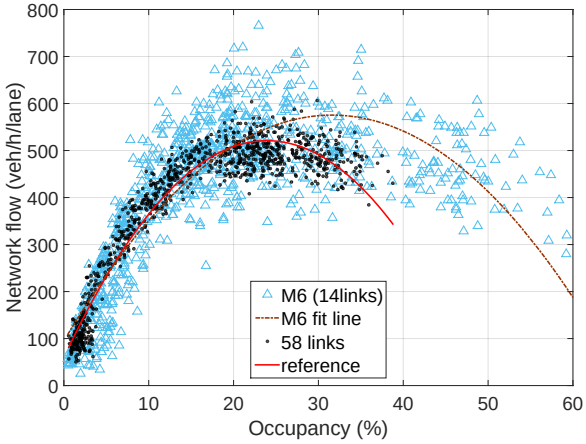
(a) Model 2



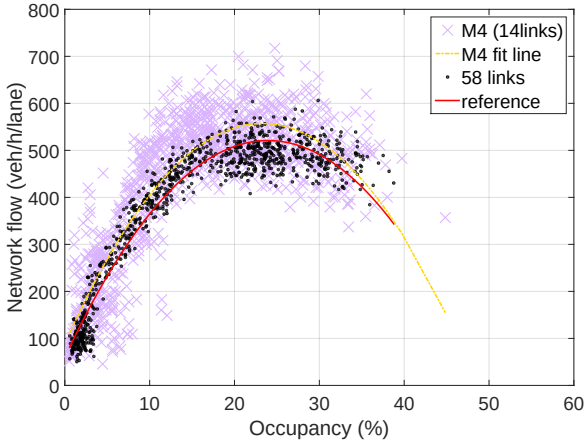
(d) Model 5



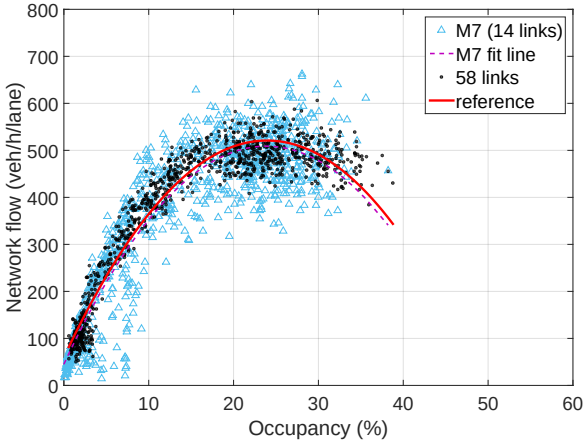
(b) Model 3



(e) Model 6

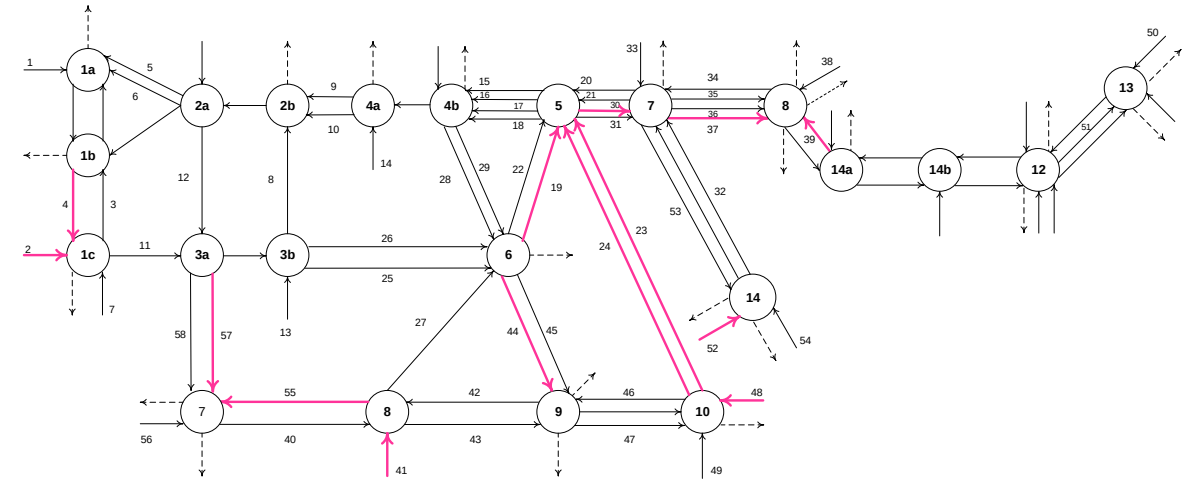


(c) Model 4

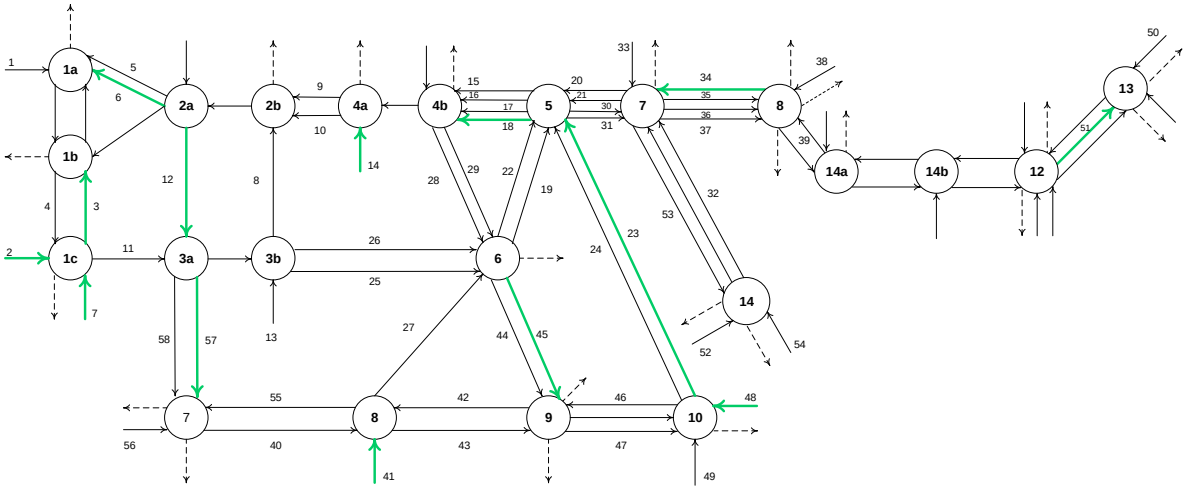


(f) Model 7

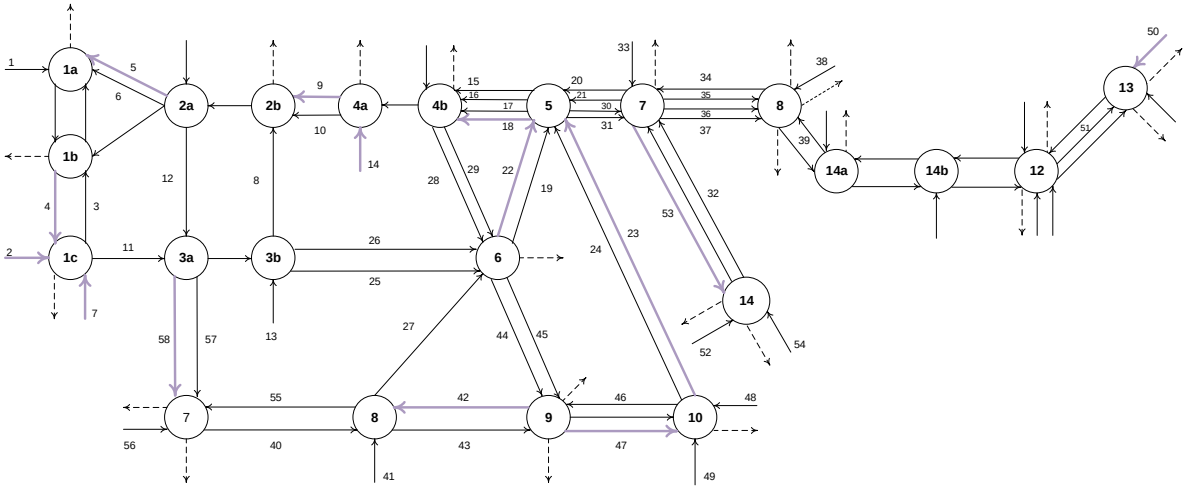
Figure 3.30: NFD of less number of links for Friday according to different models.



(a) Link (sensor) selection of Model 2



(b) Link (sensor) selection of Model 3



(c) Link (sensor) selection of Model 4

Figure 3.31: Selected link (sensor) locations according to different models.

3.4 Discussion

This chapter proposed an information theoretic-based framework for the efficient model selection and construction of sparse-measurement network-wide fundamental diagrams. The approximated sparse-measurement fundamental diagrams, which are in principle less costly in terms of infrastructure requirements, can be used for the efficient monitoring and control of congested urban areas. For the optimal selection of sensors a set covering GIP problem was developed. A measure of correlation between random variables was introduced as a "distance" metric to provide sufficient coverage and information accuracy. The developed GIP is data-driven, including data sources of information from traffic occupancies, flows, and the spatial distance between neighbour sensors. In this way different features of the full-measurements NFD can be preserved. Based on the different available data sources of information, four rival GIP models are considered. Two ad-hoc methods namely blind and radius strategy were also considered for comparison. The Kullback-Leibler divergence was used to measure the dissimilarity between probability mass functions corresponding to different models obtained from the solution of the GIP problem.

Results from the application of the proposed framework to empirical data from around sixty sensors have demonstrated that the obtained sparse-measurement rival diagrams were able to preserve the shape and main features of the operational full-measurement diagram. To assess the performance of the different models a number of qualitative and quantitative criteria were used, including the measure of correlation, the optimal cost of the GIP, and the Kullback-Liebler divergence. The optimal cost (number of sensors selected) of the GIP models is decreased with the value of correlation while Kullback-Liebler divergence is increased, indicating information loss when number of sensors is decreased.

The GIP-based models indicated a good fit with the reference NFD and preserved the critical occupancy between 20% and 30%. On the other hand, models based on the radius strategy (including the same number of selected sensors for comparative results) alone indicated the worst approximation. This confirms our initial conjecture and motivation to pursue this work, that the spatial distance cannot be employed as a single metric but can be combined with other sources of information (as in the developed GIP models) given that sensors or links located (far one each other) at arbitrary areas of the network can provide similar information and thus should be excluded. The blind strategy indicated an acceptable fit though the capacity flow exceeded the acceptable error. It is observed that there were many solutions of rival models (GIP-based or blind-based) with close Kulllback-Liebler divergence value. This could be attributed to the size and topology of network used in this study. The availability of real empirical data for this network though gave us some good evidence on the approximation problem and performance of different models. It is expected that a larger heterogeneous network with multiple centers of congestion would make the performance of blind or other ad-hoc strategies to deteriorate.

Chapter 4

Multi-gated perimeter flow control

4.1 Introduction

The idea of perimeter control using the concept of NFD was discussed in Chapter 2. Existing studies assume that the input (gated) flows obtained from a perimeter control strategy are equally (or not optimally) distributed to a number of candidate junctions at the periphery of the network, without taking into account different geometric characteristics (e.g., cycle times, saturation flows) of origin links and other state and control constraints. This chapter addresses this problem by designing efficient control schemes for the optimal distribution of input flows across the perimeter of single regions or the boundaries of neighbour regions which allows for the consideration of geometric characteristics at individual gated links, and state and control constraints. I is referred as *multi-gated perimeter flow control* (MGC) scheme.

In the proposed multi-gated perimeter flow control scheme, the NFD is utilised to describe the traffic dynamics within the protected area. For the exterior of the protected area, the basic state space model is expanded with additional state variables account for queues at store-and-forward origin links at the periphery. In this way, limited origin links storage capacity is respected, hence interference with adjacent street traffic outside of the protected network area will be avoided. The integrated model is then used to formulate a convenient convex or nonlinear optimal control problem with constrained control and state variables. Model-based predictive control (MPC) in a rolling horizon framework is used to solve the corresponding constrained optimal control problem. The optimal control problem aims to balance the relative queues at origin links and maintain the vehicle accumulation in the protected network around a desired point, while the system's throughput is maximised.

This chapter also presents practical flow allocation policies for single-region perimeter control strategies without explicitly considering entrance link dynamics. These strategies employ a two-step procedure. Firstly, an ordered flow is obtained from an unconstrained controller that does not directly incorporate the operational constraints into the controller synthesis. Secondly, the ordered-flow is then distributed to equivalent entrance link green stages at the perimeter with

the help of a flow allocation policy. This chapter offers two perimeter-ordered flow allocation policies, namely *capacity-based flow allocation policy* (CAP) and *optimisation-based flow allocation policy* (OAP), to facilitate the real-time deployment of such strategies. Further, these flow allocation policies are benchmarked against the proposed multi-gated perimeter flow control in a protected area of Downtown San Francisco, CA, including 110 junctions, 440 links, and 15 controlled gates. The obtained results demonstrate the efficiency and equity properties of the proposed multi-gated scheme to better manage excessive queues outside of the protected network area and optimally distribute the input flows.

The main contributions and features of the proposed multi-gated perimeter flow control scheme are as follows: (a) it offers in the literature an efficient scheme to determine the optimal distribution of input flows for a number of gates located at the periphery of a protected network area with respect to link's storage capacities and geometric characteristics; (b) it fulfils the requirement of equity for drivers using different gates to enter a protected network area; (c) it manages excessive queues outside of the protected network area, resulting from the perimeter control; (d) it proposes for the first time in the relevant literature practical flow allocation policies for single-region perimeter control strategies.

The rest of this chapter is organised as follows. Section 4.2 describes the traffic dynamics of single region networks. Section 4.3 presents the proposed MPC-based multi-gated perimeter flow control strategy. Section 4.4 offers practical perimeter-ordered flow allocation policies. Section 4.5 provides a case study of the application of the proposed MGC and allocation policies CAP & OAP to the protected area of Downtown San Francisco, CA. Section 4.6 summarises the chapter. To assist the reader, some background on the MPC is provided in Appendix B.

4.2 Traffic dynamics of single-region networks

4.2.1 Dynamics within the protected area

This section discusses the traffic dynamics of a single-region network without considering entrance link dynamics. Consider a protected network of a single region. The protected network with its inner traffic dynamic can be treated macroscopically as a single-region dynamic system with vehicle accumulation $n(t)$, $t \geq 0$ as a single state variable [23]. To this end, assume there exists a well-defined function $O(n(t))$ (veh/h) that provides the estimated rate flow (output) at which vehicles complete trips per unit time either because they finish their trip within the network or because they move outside of the network. This function describes steady-state behaviour of single-region homogeneous networks if the input to output dynamics are not instantaneous and any delays are comparable with the average travel time across the network. The output (throughput) function $O(n(t))$ of a network can be easily determined if trip completion rates or Origin-Destination (OD) data are available in real-time (e.g. from vehicles equipped

with GPS trackers capable of providing locational data at any given time). Alternatively, the output can be expressed as $O(n(t)) = (l/L)O_c(n(t))$, where L (m) is the average trip length in the network, l (m) is the average link length, and O_c (veh/h) is the total network circulating flow. In general, the circulating flow O_c can be estimated by Edie's generalised definition of flow if $n(t)$ is observed in real-time, i.e. weighted average of link flows with link lengths [29].

Let $q_{\text{in}}(t)$ (veh) be the inflow to the region at time t. Also, let $q_{\text{out}}(t)$ (veh/h) and $d_n(t)$ (veh/h) be the outflow and the uncontrolled traffic demand (disturbances) of the protected network at time t, respectively. For a single input variable, $d_n(t)$ is internal (off-street parking for taxis and pockets for private vehicles) non-controlled inflows. The dynamics of the system are governed by the following nonlinear conservation equation

$$\dot{n}(t) = q_{\text{in}}(t - \tau_n) - q_{\text{out}}(t) + d_n(t), \tag{4.1}$$

where $q_{\text{out}}(t)$ is in general a nonlinear function of vehicle accumulation $n(t)$ and τ_n is the travel time needed for vehicles to approach the protected network area. Without loss of generality, it is assumed that $\tau_n = 0$. Since the system evolves slowly with time t, it may assumed that outflow $q_{\text{out}}(t) \propto O_c(n(t)$, and it may thus be given in terms of the output $O(n(t))$. Note that $q_{\text{in}}(t)$ are the input variables of the controlled region, to be calculated by a perimeter flow control strategy.

Now consider a protected network with a number of candidate gates controlling the allowable flow entering the protected area. Let $\mathcal{O} = \{1, 2, \ldots\}$ be the set of controlled gates located at its periphery. The set $\mathcal{O}$ includes all the origin links whose outflow is essentially entering into the protected network from a number of controlled gates (e.g. signalised junctions or toll stations). Hence the total inflow to the network is given by $q_{in} = \sum_{o=1}^{|\mathcal{O}|} q_o$.

The dynamics of the system governed are still the same as in (4.1) with additional control variables and nonlinear conservation equation as follows,

$$\dot{n}(t) = \sum_{o=1}^{|\mathcal{O}|} q_o(t - \tau_o) - q_{\text{out}}(t) + d_n(t), \tag{4.2}$$

where $q_{\text{out}}(t)$ is in general a nonlinear function of vehicle accumulation $n(t)$ and τ_o is the travel time needed for vehicles to approach the protected network area from origin link $o \in \mathcal{O}$. The time lags τ_o may be translated into an according number of time steps for a discrete-time representation, provided a closed system with inflows q_o and outflow q_{out}. Without loss of generality, $\tau_o = 0, \forall\ o \in \mathcal{O}$ is assumed, i.e., vehicles released from the controlled gates can immediately get access to the protected network. Moreover, since the system evolves slowly with time t, one may assume that outflow $q_{\text{out}}(t) \propto O_c(n(t)$, and it may thus be given in terms of the output $O(n(t))$. Note that $q_o(t)$, $o \in \mathcal{O}$ are the input variables of the controlled gates/entrances, to be calculated by a mutli-gated perimeter flow control strategy.

The input flow from the perimeter control strategy q_o can be then converted to feasible en-

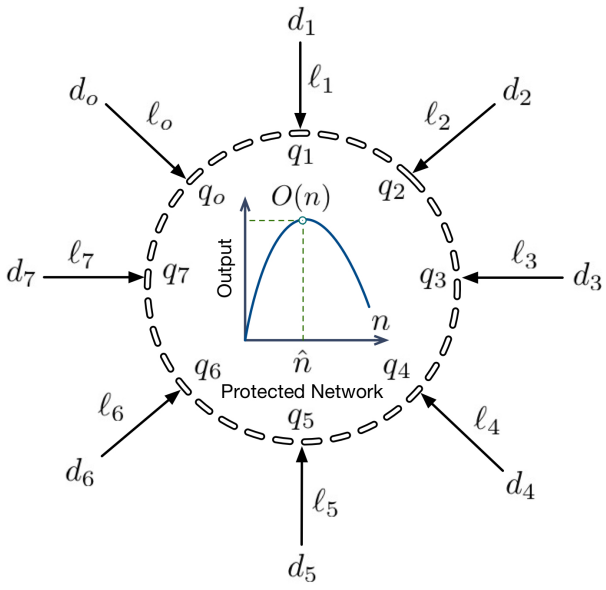


Figure 4.1: Protected network with entrance link dynamics.

trance link green time g_o for real life implementation in a traffic signal plan. The transformation into green time is given by:

$$g_o = \frac{c_o \times q_o}{s_o} \tag{4.3}$$

where g_o is the green time, s_o and c_o are saturation and cycle time at gating link o.

4.2.2 Entrance link dynamics

This section discusses the traffic dynamics of a single-region network with entrance link dynamics. In principle, the origin links at the periphery of the protected network would have different geometric characteristics such as length, number of lanes, capacity and saturation flows. For this reason, the original model is enhanced by considering dynamics outside of the protected area. Hence in this section dynamics of both outside and inside controlled areas are considered.

Consider a protected network area with number of controlled gates $o \in \mathcal{O} = \{1, 2, \dots\}$ located at its periphery as shown in Figure 4.1. The set $\mathcal{O}$ includes all the origin links whose outflow is essentially entering into the protected network from number of controlled gates/entrances. Example of controlled gates could be signalised junctions or toll stations.

The nonlinear conservation equation for the dynamics inside the protected network are is the same as in (4.2). To describe traffic dynamics outside of the protected area, the basic state-space model (4.2) is augmented with additional state variables for the queues at store-and-forward entrance links at the periphery. Each origin link ℓ_o receives traffic demand d_o and forward it into the protected network, as shown in Figure 4.1. The queuing model for the entrance link

dynamics is described by the following conservation equation

$$\dot{\ell_o}(t) = d_o(t) - q_o(t), \quad o \in \mathcal{O} = \{1, 2, \ldots\} \tag{4.4}$$

where $\ell_o(t)$ (veh) and $d_o(t)$ (veh/h) are the vehicle queue and traffic demand in origin link o at time t, respectively.

The integrated model (4.2), (4.4) can be extended to consider a broader class of state and control constraints. For example, inequality state and control constraints may be introduced to preserve congested phenomena within the protected network and to avoid long queues and delays at the perimeter of the network where gating is literally applied. These constraints may be brought to the form

$$\begin{aligned} &0 \leq n(t) \leq n_{\max} \\ &0 \leq \ell_o(t) \leq \ell_{o,\max}, \quad o \in \mathcal{O} = \{1, 2, \ldots\} \\ &q_{o,\min} \leq q_o(t) \leq q_{o,\max}, \quad o \in \mathcal{O} = \{1, 2, \ldots\} \end{aligned} \tag{4.5}$$

where $n_{\max}$ is the maximum vehicle accumulation of the protected network; $\ell_{o,\max}$ is the maximum permissible capacity of link $o \in \mathcal{O}$; $q_{o,\min}$, $q_{o,\max}$ are the minimum and maximum permissible outflows, respectively; and, $q_{o,\min} > 0$ to avoid long queues and delays at the periphery of the network. Link capacities and maximum vehicle accumulation depend on geometric characteristics of the origin links (length, number of lanes) and the topology of the protected network, respectively. Minimum and maximum permissible outflows can easily be determined given saturation flows, minimum and maximum green times, and cycle times of a nominal traffic signal plan (or corresponding toll ticket) at each controlled gate of the protected network.

State equation (4.4) is linear, though a more accurate nonlinear form can be written to account storage capacity and dispersion of the flow phenomena within store-and-forward origin links. In this case, the outflow function of each gate $o \in \mathcal{O}$ is given by

$$q_o(t) = \begin{cases} 0, & \text{if } n(t) \geq c n_{\max} \\ \min\{d_o(t), \tilde{q}_o(t)\}, & \text{otherwise} \end{cases}, \tag{4.6}$$

where $\tilde{q}_o(t)$, $o \in \mathcal{O}$, are now the input variables to be calculated by a multi-gated perimeter flow control strategy; and, $c \in (0.5, 1)$ is a scalar introduced to prevent overflow phenomena within the protected network area. Note that, when using (4.6), the state constraints in (4.5) for all origin links ℓ_o, $o \in \mathcal{O}$ are considered indirectly and may hence be dropped; indeed the gated outflow in (4.6) becomes zero if there is no queue in the corresponding origin link or if the protected network is oversaturated (determined by c).

4.2.3 State-space model

The presented model can be viewed as a nonlinear process with input variables ($m = |\mathcal{O}|$ denotes the cardinality of $\mathcal{O}$) $\mathbf{u}^\mathsf{T} = \begin{bmatrix} q_1 & q_2 & \cdots & q_{|O|} \end{bmatrix} \in \mathbb{R}^m$, state variables $\mathbf{x}^\mathsf{T} = \begin{bmatrix} n & \ell_1 & \ell_2 & \cdots & \ell_{|O|} \end{bmatrix} \in \mathbb{R}^{m+1}$, and disturbances $\mathbf{d}^\mathsf{T} = \begin{bmatrix} d_n & d_1 & d_2 & \cdots & d_{|O|} \end{bmatrix} \in \mathbb{R}^{m+1}$. Then, the continuous-time nonlinear state system (4.2), (4.4), (4.6) with constraints (4.5) for a protected network with controlled gates $o \in \mathcal{O}$, may be rewritten in compact vector form following as in [1],

$$\dot{\mathbf{x}}(t) = \mathbf{f}\left[\mathbf{x}(t), \mathbf{u}(t), \mathbf{d}(t), t\right], \; t \geq 0, \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{4.7}$$

$$\mathbf{0} \leq \mathbf{x}(t) \leq \mathbf{x}_{\max}, \tag{4.8}$$

$$\mathbf{u}_{\min} \leq \mathbf{u}(t) \leq \mathbf{u}_{\max}, \tag{4.9}$$

where $\mathbf{f}$ is a nonlinear vector function reflecting the right-hand side of (4.2), (4.4); $\mathbf{x}_0$ is a known initial state; and $\mathbf{x}_{\max}$, $\mathbf{u}_{\min}$, $\mathbf{u}_{\max}$ are vectors of appropriate dimension reflecting the upper and lower bounds of constraints (4.5).

Assuming a nonlinear representation of $q_{\text{out}}(t) \triangleq O(n(t))$, the continuous-time nonlinear model (4.7) may be linearised around some set point $\hat{\mathbf{s}}^\mathsf{T} = \begin{bmatrix} \hat{\mathbf{x}}^\mathsf{T} & \hat{\mathbf{u}}^\mathsf{T} & \hat{\mathbf{d}}^\mathsf{T} \end{bmatrix}$, and directly translated into discrete-time, using Euler first-order time discretisation with sample time T, as follows

$$\Delta\mathbf{x}(k+1) = \mathbf{A}\Delta\mathbf{x}(k) + \mathbf{B}\Delta\mathbf{u}(k) + \mathbf{C}\Delta\mathbf{d}(k) \tag{4.10}$$

where $k = 0, 1, \ldots, N_o - 1$ is a discrete time index with optimisation horizon N_o; $\Delta(\cdot) \triangleq (\cdot) - \hat{\cdot}$ for all vectors; and $\mathbf{A} = \partial\mathbf{f}/\partial\mathbf{x}|_{\hat{\mathbf{s}}}$, $\mathbf{B} = \partial\mathbf{f}/\partial\mathbf{u}|_{\hat{\mathbf{s}}}$, $\mathbf{C} = \partial\mathbf{f}/\partial\mathbf{d}|_{\hat{\mathbf{s}}}$ are the state, control, and disturbance matrices, respectively (see section B.1 in Appendix B). This discrete-time linear model is completely controllable and reachable, and will be used as a basis for control design.

The sample time interval T is literally selected to be a common multiple of cycle lengths of all controlled gates at the periphery of the protected network, while $T \in [3, 5]$ minutes is usually appropriate for constructing a well-defined outflow function $O(n(t))$, given experimental data. In principle, origin link dynamics (4.4) are much faster than the dynamics of the protected network (4.2) (governed by the network fundamental diagram, which evolves slowly in time). Therefore two different time steps T_n and T_ℓ (where $T_\ell \ll T_n$) can be employed for (4.2) and (4.4), respectively, to account storage capacity and dispersion of the flow phenomena within store-and-forward origin links; and thus increase model accuracy. By introducing different time steps, the state variables for origin links ℓ_o, $o \in \mathcal{O}$, are allowed to change their value more frequently than the state of the protected network n and control variables q_o, $o \in \mathcal{O}$.

The state-space matrices of the augmented model (4.10) with dynamics outside the protected area are: state matrix $\mathbf{A} = \text{diag}(1 - O'(\hat{n}(k)) \times T, 1, \ldots, 1) \in \mathbb{R}^{(m+1)\times(m+1)}$; control

matrix $\mathbf{B} = T \begin{bmatrix} \mathbf{1}_{1\times m} & -\mathbf{I}_{m\times m} \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{(m+1)\times m}$; and, disturbance matrix $\mathbf{C} = \mathrm{diag}(T, \ldots, T) \in \mathbb{R}^{(m+1)\times(m+1)}$.

4.3 Multi-gated perimeter flow control via MPC

This section presents the perimeter control objection and solution of the multi-gated perimeter flow control via MPC (see Appendix B for more information).

4.3.1 Control objective

A natural control objective for the traffic system considered is to minimise the total time that vehicles spend in the system including both time waiting at origin links to enter and time traveling in the protected network. Actually the minimisation of the total time spent is equivalent to the maximisation of the total exit flow (or trip completion rate) from the protected network, under the assumption of given control-independent demand inflows and of infinitely long origin link queues (unconstrained vehicle storage). However, such a policy may induce unbalanced gating of vehicles at the origin links of the protected network, and, as a consequence, would lead to long queues and overflow phenomena within origin links. Unbalanced gating would also violate the requirement of equity for drivers using different gates to enter a protected network area.

Given these observations, a suitable control objective for a protected network area with origin links queue dynamics aims at: (a) equalising the relative vehicle queues $\ell_o/\ell_{o,\max}$, $o \in \mathcal{O}$ over time, and (b) maintaining the vehicle accumulation in the protected network around a set (desired) point $\hat{n}$ while the system's throughput is maximised. A quadratic criterion that considers this control objective has the form

$$J = \frac{1}{2} \sum_{k=0}^{N_o-1} \left(\|\Delta \mathbf{x}(k)\|^2_{\mathbf{Q}} + \|\Delta \mathbf{u}(k)\|^2_{\mathbf{R}} \right) \tag{4.11}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are positive semi-definite and positive definite diagonal weighting matrices, respectively. The diagonal elements of $\mathbf{Q}$ (see definition of vector $\mathbf{x}$ in previous section) are responsible for balancing the relative vehicle accumulation of the protected network $n/n_{\max}$ and the relative vehicle queues $\ell_o/\ell_{o,\max}$, $o \in \mathcal{O}$. Given that vehicle storage in the protected network is significantly higher than in the origin links, a meticulous selection of diagonal elements is required. A practicable choice is to set $\mathbf{Q} = \mathrm{diag}(1/w, 1/\ell_{1,\max}, \ldots, 1/\ell_{|\mathcal{O}|,\max})$, where the scale of $w \ll n_{\max}$ is of the order of $\sum_{o=1}^{|O|} \ell_{o,\max}$ to achieve equity (see also [2, 27]). It becomes quite clear here that equity at origin links and efficiency of the protected network area are partially competitive criteria, hence a perimeter flow control strategy should be flexible enough to accommodate a particular trade-off (i.e. to give priority to the protected network or the outside area, e.g. to manage better excessive queues) to be decided by the responsible network authorities.

Finally, the choice of the weighting matrix $\mathbf{R} \triangleq r\mathbf{I}$, $r > 0$ can influence the magnitude of the control actions and thus r should be selected via a trial-and-error process.

4.3.2 Model-predictive control

Model-based predictive control is a repetitive optimisation scheme, where at each time step an open-loop optimal control problem with finite horizon N_o and predicted demands $\mathbf{d}(k)$ over a prediction horizon N_p is optimised, then only the first control move is applied to the plant and the procedure is carried out again. This rolling-horizon procedure closes the loop that is avoids myopic control actions while embedding a dynamic open-loop optimisation problem in a responsive environment. Predicted demand flows $\mathbf{d}(k)$ may be calculated by use of historical information or suitable extrapolation methods.

Given the known initial state $\mathbf{x}(0) = \mathbf{x}_0$, a static convex optimisation problem may be formulated over N_o due to the discrete-time nature of the involved process. To see this, assume $N_o = N_p$ and define the vectors

$$\begin{aligned}
\Delta\mathbf{X} &= \begin{bmatrix} \Delta\mathbf{x}(1)^\mathsf{T} & \Delta\mathbf{x}(2)^\mathsf{T} & \cdots & \Delta\mathbf{x}(N_o)^\mathsf{T} \end{bmatrix}^\mathsf{T} \\
\Delta\mathbf{U} &= \begin{bmatrix} \Delta\mathbf{u}(0)^\mathsf{T} & \Delta\mathbf{u}(1)^\mathsf{T} & \cdots & \Delta\mathbf{u}(N_o-1)^\mathsf{T} \end{bmatrix}^\mathsf{T} \\
\Delta\mathbf{D} &= \begin{bmatrix} \Delta\mathbf{d}(0)^\mathsf{T} & \Delta\mathbf{d}(1)^\mathsf{T} & \cdots & \Delta\mathbf{d}(N_p-1)^\mathsf{T} \end{bmatrix}^\mathsf{T}.
\end{aligned}$$

Assuming now availability of demand flow predictions at the origin links of the protected network over a prediction horizon N_p, i.e. $\Delta\mathbf{d}(k) \neq \mathbf{0}$, $k = 0, 1, \ldots, N_p - 1$, minimisation of the performance criterion (4.11) subject to (4.10) leads to the analytical solution:

$$\Delta\mathbf{U} = -\mathbf{H}^{-1}\mathbf{F}[\mathbf{x}(0) + \mathbf{G}\Delta\mathbf{D}], \tag{4.12}$$

where $\mathbf{H} = \mathbf{\Gamma}^\mathsf{T}\mathcal{Q}\mathbf{\Gamma} + \mathcal{R}$ is the Hessian of the corresponding quadratic program (QP), $\mathbf{F} = \mathbf{\Gamma}^\mathsf{T}\mathcal{Q}\mathbf{\Omega}$, and $\mathbf{G} = \mathbf{\Gamma}^\mathsf{T}\mathcal{Z}$. The matrices $\mathbf{\Gamma}$ and $\mathbf{\Omega}$ may be readily specified from the integration of (4.10) starting from the initial state $\mathbf{x}(0)$. The weighting matrices $\mathcal{Q}$, $\mathcal{R}$, $\mathcal{Z}$ are as follows,

$$\begin{aligned}
\mathcal{Q} &= \mathrm{diag}(\mathbf{Q}, \mathbf{Q}, \cdots, \mathbf{Q}), \\
\mathcal{R} &= \mathrm{diag}(\mathbf{R}, \mathbf{R}, \cdots, \mathbf{R}), \\
\mathcal{Z} &= \mathrm{diag}(\mathbf{Q}, \mathbf{Q}, \cdots, \mathbf{Q}).
\end{aligned}$$

Given that $\mathbf{R} \succ \mathbf{0}$ in the cost criterion (4.11) the Hessian $\mathbf{H}$ is positive definite, and thus the QP is convex and has a global optimum. Note that the third term may be regarded as a feedforward term, accounting for future disturbances. Clearly for $N_o \to \infty$ and vanishing

disturbances, i.e., $\Delta\mathbf{d}(k) = \mathbf{0}$, $k = 0, 1, \ldots, N_p - 1$, a Linear-Quadratic or a Linear-Quadratic-Integral regulator may be derived as in [1].

Using the above formalism, the problem of minimising (4.11) subject to the equality constraints (4.10) and inequality constraints (4.8)–(4.9) is express as follows:

$$\begin{aligned} &\min_{\mathbf{U}} \frac{1}{2}\mathbf{U}^{\mathsf{T}}\mathbf{H}\mathbf{U} + \mathbf{U}^{\mathsf{T}}[\mathbf{F}\mathbf{x}(0) - \mathbf{H}\hat{\mathbf{U}} + \mathbf{G}\Delta\mathbf{D}] \\ &\text{subject to:} \\ &\mathbf{L}\mathbf{U} \leq \mathbf{W} \end{aligned} \tag{4.13}$$

where $\mathbf{L}$ and $\mathbf{W}$ are matrices reflecting the lower and upper bounds of the state and control constraints (given state integration starting from the initial state $\mathbf{x}_0$) over the optimisation horizon N_o (see Appendix B for details). Once the open-loop QP problem (B.28) is solved from the known initial $\mathbf{x}(0)$ and predicted disturbances $\mathbf{d}(k), k = 0, 1, \ldots, N_p - 1$, the rolling horizon scheme applies, at the current time k, only the first control move, formed by the first m components of the optimal vector $\mathbf{U}^*(\mathbf{x}_0)$ in (B.28). This yields a control law of the form

$$\mathbf{u}(k) = \mathcal{M}[\mathbf{x}(k), \mathbf{d}(\kappa)], \quad \kappa = k, k+1, \ldots, k+N_p-1 \tag{4.14}$$

where $\mathbf{x}(k) = \mathbf{x}_0$, $k = 0, \ldots, N_o - 1$ is the current state of the system and $\mathcal{M}$ is a linear mapping from the state and disturbance spaces to control. Then the whole procedure is repeated at the next time instant, with the optimisation horizon kept constant. Note that the analytical solution (4.12) for the *unconstrained problem* is of particular interest; given that the optimal solution for the constrained problem has a similar form in a region of the state space where the state of the system vanishes, i.e. $\Delta\mathbf{x} = \mathbf{0}$ or $\mathbf{x} = \hat{\mathbf{x}}$.

4.4 Practical perimeter-ordered flow allocation policies

4.4.1 Introduction

This section presents practical flow allocation policies for single-region perimeter control strategies without explicitly considering entrance link dynamics. The problem under consideration is to allocate a global perimeter-ordered flow to a number of candidate gates/junctions at the periphery of the network by taking into account the different geometric characteristics of origin links, i.e., length, number of lanes, storage capacity, etc. The global flow $q_{\mathrm{G}}(k) = \sum_{o=1}^{|\mathcal{O}|} q_o(k)$ at discrete time k can be ordered by any perimeter control strategy. For instance, a single-input single-output controller with only state equation (4.2) and cost criterion (4.11), or the bang-bang policy in [23], or the feedback controllers in [1, 55], or other similar strategies. Previous perimeter flow control strategies without explicitly considering entrance link dynamics (e.g. [1, 9, 55]),

assume that a single-ordered input flow is distributed in a non-optimal or a posteriori way (e.g., equally distributed or with respect to saturation flows) to a number of candidate gates at the periphery of a protected networks area. These strategies employ a two-step procedure. Firstly, an ordered flow is obtained from an unconstrained controller that does not directly incorporate the operational constraints into the controller synthesis. Secondly, the ordered-flow is then distributed to equivalent entrance link green stages at the perimeter with the help of a flow allocation policy. In the sequel, two perimeter-ordered flow allocation policies is proposed to facilitate the real-time deployment of such strategies. These flow allocation policies are later benchmarked against the proposed multi-gated perimeter flow control in Section 4.5.3.4.

4.4.2 Capacity-based flow allocation policy (CAP)

In principle, the distribution of the global perimeter-ordered flow among the controlled gates should be according to appropriately predefined portions. These portions are typically proportional to the links' vehicle storage capacities or nominal flows. In other words, links with high storage capacity or high nominal flows will carry more flow. Moreover, drivers waiting in gated links with similar storage capacity to enter the network from the periphery would potentially experience similar delays. In this section, a capacity-based allocation policy is proposed, given that storage vehicle capacity is more important than the nominal saturation flow of each link, particularly under strong gating. In a similar vein, one can develop an allocation policy based on the nominal flows.

The proposed Capacity-based flow Allocation Policy (CAP), distributes the prevailing global perimeter-ordered flow $q_{\mathrm{G}}(k)$ to the controlled gates $o \in \mathcal{O} = \{1, 2, \ldots\}$ according to the link storage capacities $\ell_{o,\max}$. The desired distribution attempted via CAP is meant to be active both during the emptying and gating phases, so that equal free relative vehicle storages are provided to each entrance link (in spite of different storage capacities) in the event of strong gating or upstream traffic demand. To start with, the capacity ratio of each entrance link is defined by:

$$r_o = \frac{\ell_{o,\max}}{\sum_{o=1}^{|\mathcal{O}|} \ell_{o,\max}}, \quad o = 1, 2, \ldots, |\mathcal{O}|. \tag{4.15}$$

Given the global prevailing perimeter-ordered flow $q_{\mathrm{G}}(k)$ at discrete time k, the individual subflows are then determined by:

$$q_o(k) = \hat{q}_o + r_o \times \left(q_{\mathrm{G}}(k) - \sum_{o=1}^{|\mathcal{O}|} \hat{q}_o \right), \quad o = 1, 2, \ldots, |\mathcal{O}|. \tag{4.16}$$

4.4.3 Optimisation-based flow allocation policy (OAP)

The Optimisation-based flow Allocation Policy (OAP) aims at minimising the relative difference between ordered input flow and nominal flow at each controlled gate. To this end, the following optimisation problem can be formulated and solved at each discrete time k, provided a global perimeter-ordered flow $q_{\mathrm{G}}(k)$:

$$\begin{aligned} \min_{q_o(k)} \quad & \frac{1}{2}\sum_{o=1}^{|\mathcal{O}|}\frac{\left[q_o(k)-\hat{q}_o\right]^2}{\hat{q}_o} \\ \text{subject to:} \quad & \sum_{o=1}^{|\mathcal{O}|} q_o(k) = q_{\mathrm{G}}(k) \\ & q_{o,\min} \le q_o(k) \le q_{o,\max}, \quad o = 1,2,\ldots,|\mathcal{O}|. \end{aligned} \tag{4.17}$$

The first constraint in (4.17) holds by definition, while in the second constraint individual $q_o(k)$ input flows are subject to minimum and maximum permissible outflows. This is a static optimisation quadratic programming problem that can be efficiently solved by commercial or public available software. It should be noted that if the minimum and/or maximum bounds are activated then part of the global perimeter-ordered flow $q_{\mathrm{G}}(k)$ will be wasted. On the other hand, if bound constraints are not activated then the optimisation yields the analytical solution:

$$\begin{bmatrix} \mathbf{q}(k) \\ \lambda \end{bmatrix} = \mathcal{A}^{\dagger}\mathcal{B}, \tag{4.18}$$

where $\mathbf{q}(k) \triangleq \mathbf{u}(k) \in \mathbb{R}^{|\mathcal{O}|}$ is the decision vector with elements the individual input flows $q_o(k)$, $o = 1,2,\ldots,|\mathcal{O}|$; λ is a Lagrange multiplier associated with the equality constraint in (4.17); $\mathcal{A}^{\dagger}$ is the pseudoinverse matrix that arises in standard minimum norm approximation problems; and $\mathcal{B}$ is a vector with elements the right-hand side constants of (4.17). Precisely, $\mathcal{A}$ and $\mathcal{B}$ are given by:

$$\mathcal{A} = \begin{bmatrix} \mathbf{I}_{m\times m} & -\hat{\mathbf{q}}_{m\times 1} \\ \mathbf{1}^{\mathsf{T}}_{m\times 1} & 0 \end{bmatrix}, \qquad \mathcal{B} = \begin{bmatrix} \hat{\mathbf{q}}_{m\times 1} \\ q_{\mathrm{G}}(k) \end{bmatrix},$$

where $\hat{\mathbf{q}} \triangleq \hat{\mathbf{u}}$ is the vector of nominal input flows and m refering to number of control variables (number of gates).

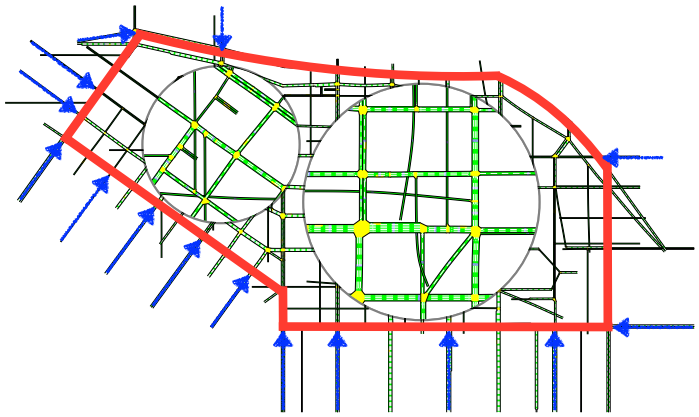


Figure 4.2: Protected network and controlled gates of entrance.

4.5 Case study and results

4.5.1 Case study description

The developed multi-gated perimeter flow control and allocation policies are applied to the 2.5 square mile area of Downtown San Francisco, CA, including 110 junctions and 440 links (see Figure 4.2). The developed approach is applied to a test network using MATLAB simulation. Although the developed models are not simulated with full network simulation, they are considered quite suitable for the demonstration and comparisons of method. Fifteen entrance links and controlled gates are illustrated with blue arrows in Figure 4.2. Figure 4.3 depicts the shape of total network circulating flow, O_c in function of $n(t)$ confirms the existence of a fundamental diagram like-shape for the study area, which shape is seen to depend on the accumulation of vehicles [1]. It can be seen that as the vehicle accumulation is increased from zero, the network flow increases to a maximum (flow capacity) and then turns down and decreases sharply to a low value possibly zero (in case of gridlock). Flow capacity (around 30×10^4 veh/h) is observed at a vehicle accumulation of about 6,000 veh. The shape of the fundamental diagram (and its critical parameters) was reproduced under different demand and OD scenarios with Dynamic Traffic Assignment activated to capture somewhat adaptive drivers in a microsimulation study via AIMSUN [1].

The shape of O_c in Figure 4.3 can be approximated by the 2nd order polynomial [1]:

$$O_c(n) = -0.0066n^2 + 87.408n \tag{4.19}$$

where $n \in [0, 12000]$. To determine the output O from O_c an average trip length $L = 1.75$ km and average link length $l = 0.25$ km were considered. The value of L is consistent with the average trip length and the travel time across the test area of San Francisco [1].

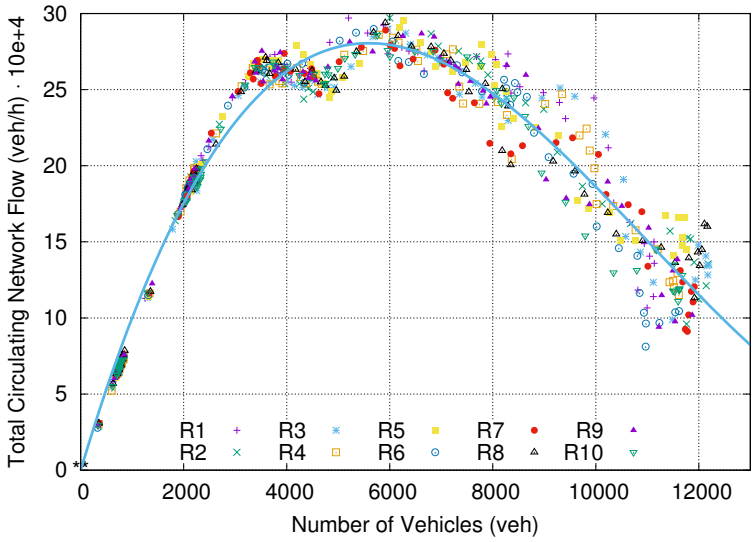


Figure 4.3: Network fundamental diagram of Downtown San Francisco [1].

4.5.2 Perimeter flow controller design

In the sequel, three different versions (Version I, II, and III) of the perimeter flow control problem are designed for comparison. All controllers designed for a single-region network but two different cases are considered. Version I concerns a single-input-single-output controller with one control variable (controlled inflow to the network) and one state variable (vehicle accumulation within the protected area). In this case a single global input flow (for the whole periphery) is calculated from the perimeter flow controller. Version II is similar to Version I but multiple input flows (control variables) apply at the perimeter of the network; and thus multiple input flows are ordered from the perimeter flow controller. Version III concerns the proposed multi-gated perimeter control (MGC), involving literally multiple input flows (control variables) with entrance link dynamics (state variables ℓ) outside of the protected network area. Version III takes into account both the different characteristics of the gates and entrance link dynamics. Table 4.2 summarises the different characteristics of the three considered perimeter flow control cases.

4.5.2.1 Geometric characteristics of gates

Table 4.1 provides the different geometric characteristics of the fifteen ($|\mathcal{O}|= 15$) entrance links and controlled gates. The third column provides the storage capacity of each controlled link that is $\begin{bmatrix} \ell_{1,\max} & \cdots & \ell_{|\mathcal{O}|,\max} \end{bmatrix} \in \mathbb{R}^{15}$. The last three columns of the table provide boundary of allowable flow as well as the reference flow at each gate that is the vectors $\mathbf{u}_{\min} = \mathbf{q}_{\min}$, $\hat{\mathbf{u}} = \hat{\mathbf{q}}$, and $\mathbf{u}_{\max} = \mathbf{q}_{\max}$, respectively. These values are calculated from the field applied signal plans presented in columns 5 (S_o: saturation flow), 6 (C: cycle length) , 7 ($g_{o,\min}$: minimum green time), 8 ($\hat{g}_o$: nominal green time), and 9 ($g_{o,\max}$: maximum green time), via gS/C. In this way, any input flows ordered by the perimeter flow control strategy are feasible traffic signal plans.

Note that traffic signals at controlled gates are all multiphase fixed-time operating on a common cycle length of 90 s for the west boundary of the area (The Embarcadero area including gates $o = 1, \ldots, 11$) and 60 s for the rest (gates $o = 12, \ldots, 15$).

4.5.2.2 Single-region control without entrance link dynamics

In the case of single-region control without entrance link dynamics (Version I) the control variable is $u = q_{\text{in}}$ and the state is $x = n$, where n is the vehicle accumulation. The state scalar A and control scalar B are constructed for the studied network on the basis of the selected $\hat{x} = \hat{n}$, $\hat{q}_{in} = 34878.12$ (veh/h), $\hat{d} = 0$ and sampling time $T = 180$ s. The desired vehicle accumulation for (4.10) (with scalar control and state) is selected $\hat{n} = 4000$. The disturbance $d = 0$ is used for the initial experiments. For the above data $A = 1 - 4.94 \times T$, $B = T$ and $C = T$. The weighting scalar $Q = 1/n_{\max}$ is selected, where $n_{\max}$ is the maximum vehicle accumulation equivalent to 12000 (veh). Three different $R = \{0.01, 0.00001, 0.0001\}$ values are tested to see the impact of control (ordered flow at the perimeter of the network) on the system.

In the case of single-region control without entrance link dynamics but multiple input flows (Version II) the control vector is $\mathbf{u} = \mathbf{q} = \begin{bmatrix} q_1, q_2, \ldots, q_m \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{15}$ and state the state variable is $x = n$. The desired steady-state is $\hat{x} = \hat{n}$, $\hat{\mathbf{u}} = \hat{\mathbf{q}} \in \mathbb{R}^{15}$, $\hat{d} = 0$ and the sampling time is $T = 180$ s. For the above data, $A = 1 - 4.94 \times T$, $\mathbf{B} = T \begin{bmatrix} \mathbf{I}_{1\times 15} \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{1\times 15}$ and $C = T$. The weighting scalar Q remains as in Version I and $\mathbf{R} = \text{diag}(r, r, \ldots, r) \in \mathbb{R}^{15\times 15}$ where $r = 0.00001$ found by trial-and-error.

4.5.2.3 Single-region control with entrance link dynamics: Multi-gated control (MGC)

In the case of the proposed multi-gated control (Version III) the desired vehicle accumulation for (4.10) is selected $\hat{n} = 4000$ veh, while $\hat{\ell}_o = 0$, $\forall o \in \mathcal{O}$. For the solution of (4.12) or (B.30) it suffices to specify the state matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, and weighting matrices $\mathbf{Q}$ and $\mathbf{R}$. All state matrices are constructed for the studied network on the basis of the selected $\hat{\mathbf{x}}^{\mathsf{T}} = [\hat{n} \quad \mathbf{0}] \in \mathbb{R}^{16}$, $\hat{\mathbf{u}} = \hat{\mathbf{q}} \in \mathbb{R}^{15}$ and $\hat{\mathbf{d}} = \mathbf{0}$, and sampling time $T = 180$ s. More precisely, $\mathbf{A} = \text{diag}(1 - 4.94 \times T, 1, \ldots, 1) \in \mathbb{R}^{16\times 16}$, $\mathbf{B} = T \begin{bmatrix} \mathbf{1}_{1\times 15} & -\mathbf{I}_{15\times 15} \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{16\times 15}$, and $\mathbf{C} = \text{diag}(T, \ldots, T) \in \mathbb{R}^{16\times 16}$. The matrix $\mathbf{Q} = \text{diag}(1/w, 1/\ell_{1,\max}, \ldots, 1/\ell_{|\mathcal{O}|,\max})$ is selected, where $w = 2000$ was found appropriate to achieve equity. The diagonal elements of $\mathbf{R}$ were set equal to $r = 0.00001$ (after a trial-and-error procedure). The disturbance vector $\mathbf{d}$ consists of the demands d_o, $o = 1, \ldots, 15$, at every origin of the protected network and disturbance d_n of the fundamental diagram. For the open-loop experiments, trapezoidal demands have been used for $d_o(k)$, $o = 1, \ldots, 15$, $k = 0, \ldots, N_p - 1$ over a predicted horizon of $N_o = N_p = 40$. To capture the uncertainty of the (scaled) fundamental diagram, particularly when the network operating in the congested regime (notice the noise for $n > 6000$ veh), d_n is selected to vary gradually

Table 4.1: Different characteristics of entrance links and controlled gates.

Gate #	Length (m)	Capacity (veh)	No lanes	Saturation Flow (veh/h)	Cycle Length (s)	Min Green (s)	Nominal Green (s)	Max Green (s)	Min Flow (veh/h)	Nominal Flow (veh/h)	Max Flow (veh/h)
1	235	128	3	5400	60	15	33	39	1350	2970	3510
2	299	109	2	3600	60	12	30	42	720	1800	2520
3	299	163	3	5400	60	15	27	39	1350	2430	3510
4	271	98	2	3600	60	12	35	42	720	2100	2520
5	261	95	2	3600	60	12	24	42	720	1440	2520
6	299	109	2	3600	60	12	30	42	720	1800	2520
7	298	109	2	3600	60	12	36	39	720	2160	2340
8	298	109	2	3600	60	12	37	42	720	2220	2520
9	296	269	5	10000	60	17	27	31	2833	4500	5167
10	296	269	5	8800	60	16	27	38	2347	3960	5573
11	299	109	2	3600	60	12	25	42	720	1500	2520
12	190	103	3	5400	90	13	42	43	780	2520	2580
13	81	44	3	5400	90	12	18	41	720	1080	2460
14	81	44	3	5400	90	12	20	41	720	1200	2460
15	341	186	3	5400	90	12	48	59	720	2880	3540

with respect to $n(k)$ in the range $[-5000, 5000]$ veh/h for $n > 6000$ veh. For the closed-loop experiments, the profile demand used is explained in Section 4.5.3.3.

4.5.2.4 Performance assessment criteria

For each control approach, two evaluation criteria are calculated; total time spent (TTS) and relative queue balance (RQB) within the protected network and outside network areas:

$$\mathrm{TTS} = T \sum_{k=0}^{N_o} \left(\sum_{o=1}^{|\mathcal{O}|} \ell_o(k) + n(k) \right) \qquad \text{(in veh} \times \text{h)} \tag{4.20}$$

$$\mathrm{RQB} = \sum_{k=0}^{N_o} \left(\sum_{o=1}^{|\mathcal{O}|} \frac{\ell_o(k)^2}{\ell_{o,\max}} + \frac{n(k)^2}{n_{\max}} \right) \qquad \text{(in veh)}, \tag{4.21}$$

where N_o is the scenario and optimisation time horizon.

Table 4.2: Summary of different controller design for perimeter flow control.

Parameter	Version I: Without entrance link dynamics (Single-input)	Version II: Without entrance link dynamics (Multi-input)	Version III: With entrance link dynamics (Multi-gated)
Desired (set point) state, $\hat{\mathbf{x}}$	$\hat{x} = [\hat{n}]$	$\hat{x} = [\hat{n}]$	$\hat{\mathbf{x}}^{\mathsf{T}} = [\hat{n} \quad \mathbf{0}]$
Desired (set point) control input, $\hat{\mathbf{u}}$	$\hat{u} = \hat{q}_{\text{G}}$	$\hat{\mathbf{u}} = \hat{\mathbf{q}}$	$\hat{\mathbf{u}} = \hat{\mathbf{q}}$
number of state	1	1	16
state variable, $\mathbf{x}$	$x = [n]$	$x = [n]$	$\mathbf{x}^{\mathsf{T}} = [n \quad \ell_1 \quad \ell_2 \dots \ell_{\mathcal{O}}]$
number of controller	1	15	15
control variable, $\mathbf{u}$	$u = q_{\text{G}}$	$\mathbf{u}^{\mathsf{T}} = [q_1 \quad q_2 \quad \dots \quad q_{\mathcal{O}}]$	$\mathbf{u}^{\mathsf{T}} = [q_1 \quad q_2 \quad \dots \quad q_{\mathcal{O}}]$
demand or disturbance, $\mathbf{d}$	$d = [d_n]$	$d = [d_n]$	$\mathbf{d} = [d_n \quad d_1 \dots d_{\mathcal{O}}]$.
state weightage, $\mathbf{Q}$	$Q = 1/n_{max}; n_{max} = 12000$ veh	$\mathbf{Q} = 1/n_{max}; n_{max} = 12000$ veh	$\mathbf{Q} = \text{diag}(1/w, 1/\ell_{1,\max}, \dots, 1/\ell_{\lvert\mathcal{O}\rvert,\max}); w = 2000$
control weightage, $\mathbf{R}$	$R = r = 0.00001$	$\mathbf{R} = r\mathbf{I}_{15\times15}; r = 0.00001$	$\mathbf{R} = r\mathbf{I}_{15\times15}; r = 0.00001$
state coefficient (matrices), $\mathbf{A}$	$A = 1 - 4.94 \times T$	$A = 1 - 4.94 \times T$	$\mathbf{A} = diag(1 - 4.94 \times T, 1, \dots, 1) \in \mathbb{R}^{16\times16}$
control coefficient (matrices), $\mathbf{B}$	$B = T$	$\mathbf{B} = T[\mathbf{I}_{1\times15}] \in \mathbb{R}^{1\times15}$	$\mathbf{B} = T[\mathbf{I}_{1\times15} \quad -\mathbf{I}_{15\times15}]^T \in \mathbb{R}^{16\times15}$
demand coefficient (matrices), $\mathbf{C}$	$C = T$	$C = T$	$\mathbf{C} = diag(T, \dots, T) \in \mathbb{R}^{16\times16}$

4.5.3 Results

4.5.3.1 Perimeter flow control without entrance link dynamics

Version I

Figure 4.4 depicts the state and control trajectories for the perimeter control problem without link dynamics, i.e. for the single-input/single-output control problem with only state equation (4.1). The model was tested with different initial states, $n(0)$ in the range $[3000, 12000]$ veh and different weights R. Here results reported on the most representative R values that indicate good control behaviour. The optimisation horizon for each simulation is 3 h (60 cycles) with zero disturbances. Vehicle accumulations $n < 4000$ veh refer to uncongested traffic condition while $n > 4000$ veh assumes congested traffic conditions.

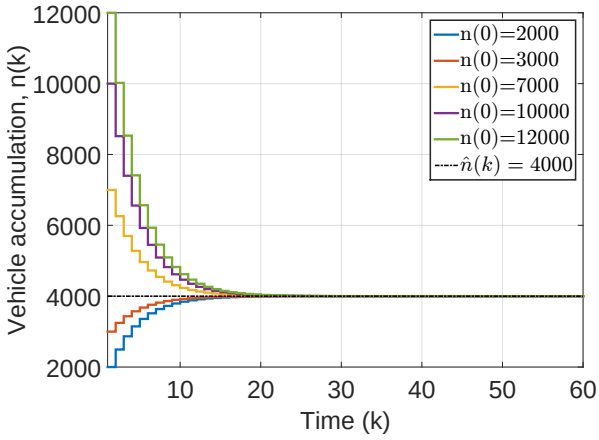
As can be seen in Figure 4.4 the strategy manages to stabilise the vehicle accumulation of the protected network around its desired state $\hat{n} = 4000$ veh starting from a number of different initial points (including the extreme case of partial gridlock when $n_{\max} = 12000$ veh). The time to stabilise the system decreases as R decreases, see Figures 4.4(a)-4.4(c). For example in Figure 4.4(c) the settling time is around $k = 12$ with $R = 0.000001$.

Figures 4.4(d)–4.4(f) demonstrate that the strategy stabilises around the desired global input flow $\hat{q}_{\text{G}} = 34878.12$ veh/h. The strategy restricts the input flow to the protected network area whenever $n > 4000$, while increases the input flows for $n < 4000$. This observation also holds when R gets small. For example, in Figure 4.4(e) where the initial accumulation is $n(0) = 10000$ veh, the ordered global input flow is decreased at around 31000 veh/h. On the other hand for $R = 10^{-6}$ (see Figure 4.4(f)), the strategy aggressively restricts the inflow to 10000 veh/h for the same initial accumulation $n(0) = 10000$ veh. Obviously, Figures 4.4(d)–4.4(f) confirm that the control is less conservative for small R. With these observations in hand, $R = 10^{-5}$ is selected as appropriate for the subsequent experiments and comparisons of different perimeter control strategies.

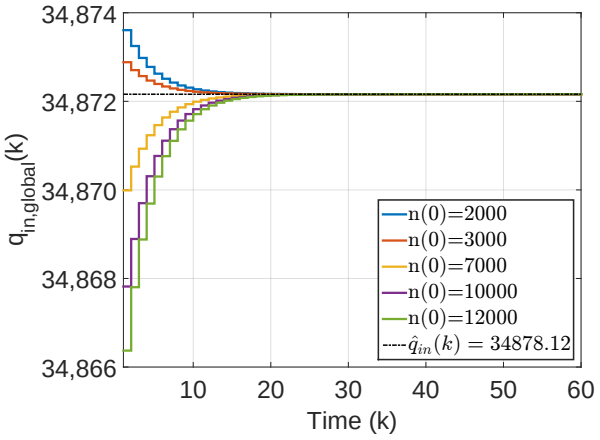
Version II

Figures 4.5–4.7 depict the state and control trajectories obtained from the application of Version II perimeter control strategy to the protected network of downtown San Francisco. The control strategy is applied for different initial states $n(0)$ in the range $[3000, 12000]$ veh and different weighing matrices $\mathbf{R} = r\mathbf{I}_{15\times 15}$ (as in Version I). Here the results reported only for $r = 0.00001$. The optimisation horizon is 3h (60 cycles) for each scenario, zero disturbance is assumed, as in Version I.

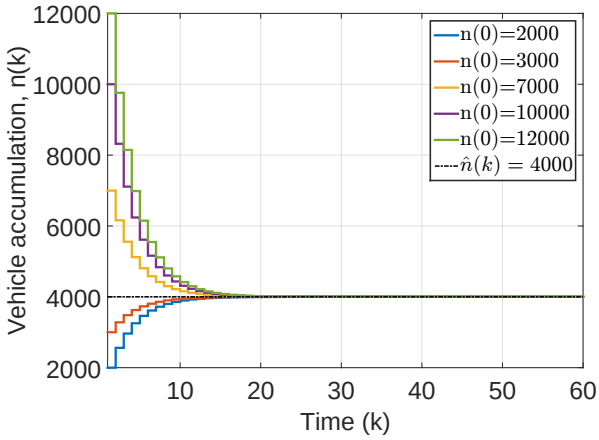
Figures 4.5–4.7 also depict the unique characteristics of each gate, e.g., different nominal input flow, different minimum and maximum flows, illustrated in green and red, respectively. As can be seen the strategy manages to stabilise all gates around their desired input flows $q_o(k)$.



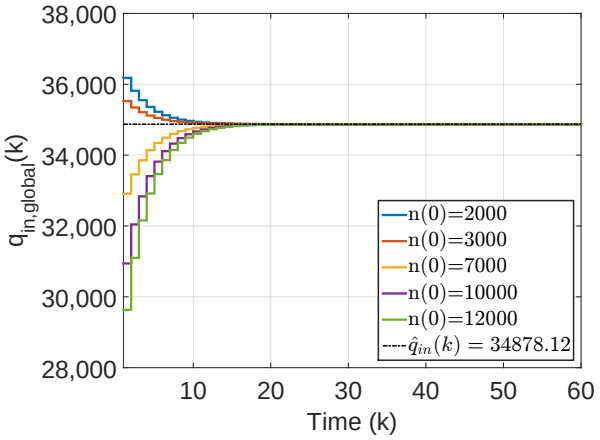
(a) $R = 10^{-2}$



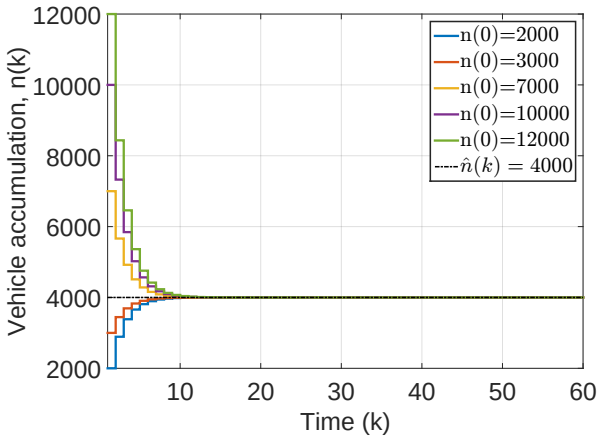
(d) $R = 10^{-2}$



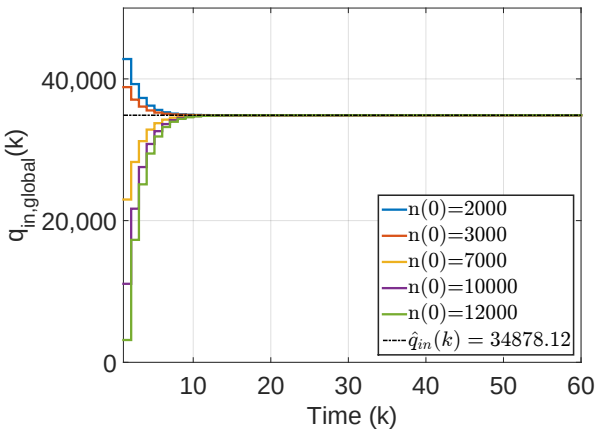
(b) $R = 10^{-5}$



(e) $R = 10^{-5}$



(c) $R = 10^{-6}$

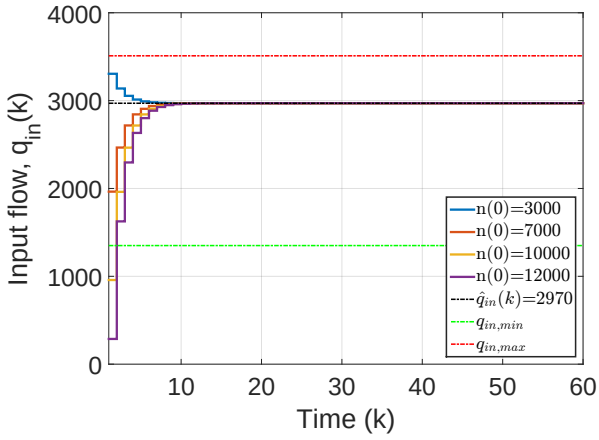


(f) $R = 10^{-6}$

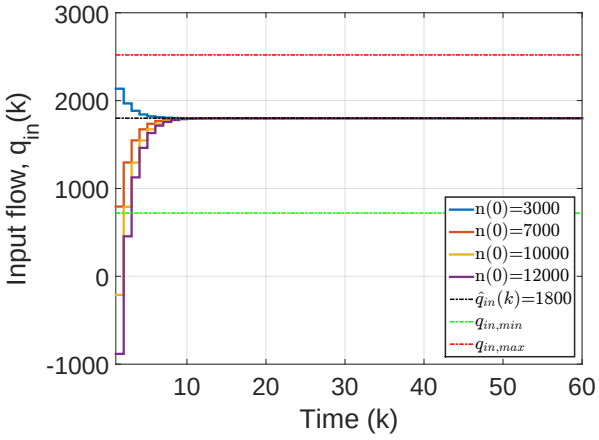
Figure 4.4: Version I: State and control trajectories for different initial states and weights.

Version II of the strategy works well for uncongested conditions, i.e., accumulation below 6000 veh. However for saturated traffic conditions, i.e., $n(0) > 6000$ veh, the ordered flows violate the lower bound input flow constraints. In other words, the strategy aggressively restricts the traffic outside of the protected network area to stabilise the overall system. On the other hand, operational constraints (minimum input flows, corresponding to minimum green times at the gated junctions) must be respected and taken into account in the optimisation; otherwise long queues will be created outside of the protected network which would lead to excessive delays and user dissatisfaction. In any case, the strategy manages (by violating the control constraints in a few congested scenarios) to stabilise the protected network around its desired state $\hat{n} = 4000$ veh (see Figure 4.7(d)).

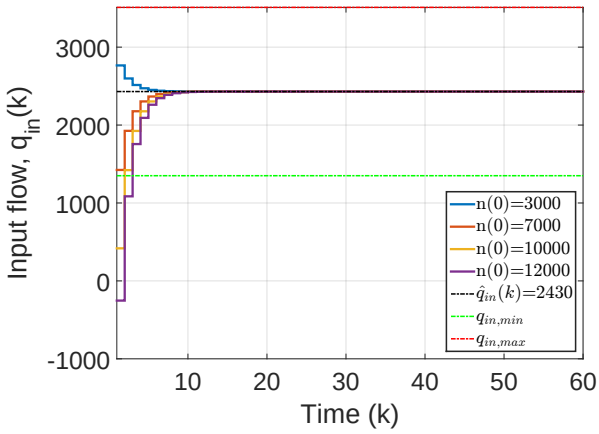
To overcome the violation of constraints, Version II of the strategy was also applied with constrained control. Figures 4.8–4.10 depict the obtained results when constrained optimisation is applied. The model was simulated with different initial states $n(0)$ in the range $[3000, 12000]$ veh and different weighing matrices $\mathbf{R} = r\mathbf{I}_{15\times 15}$ (as in Version I). Here the results reported only for $r = 0.00001$. The optimisation horizon for each simulation is 3h (60 cycles), zero disturbance is assumed, as in Version I. Version II of the strategy with constrained optimisation respects the control constraints, though it takes longer time to stabilise the overall system compared to the unconstrained problem. As can be seen in Figures 4.7(d) and 4.10(d), the strategy manages to stabilise around its desired accumulation $\hat{n} = 4000$ veh at $k > 10$, while in the unconstrained control at $k < 10$. This is the price needed to pay in order to satisfy the operational constraints.



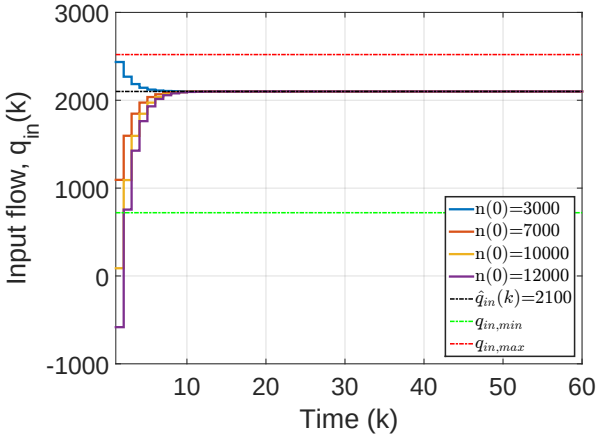
(a) Gate 1



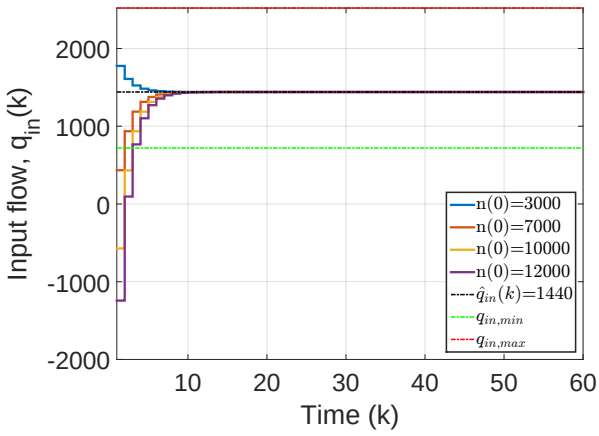
(b) Gate 2



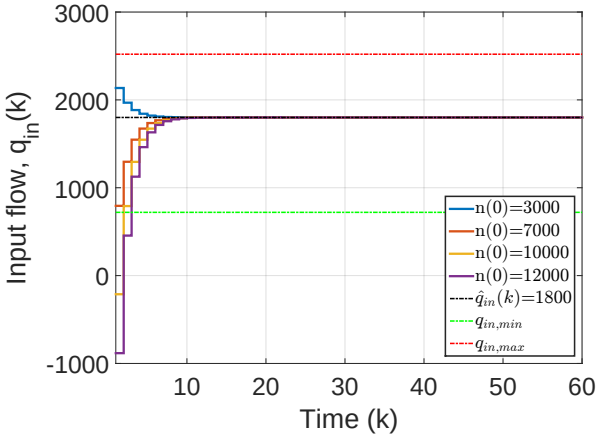
(c) Gate 3



(d) Gate 4

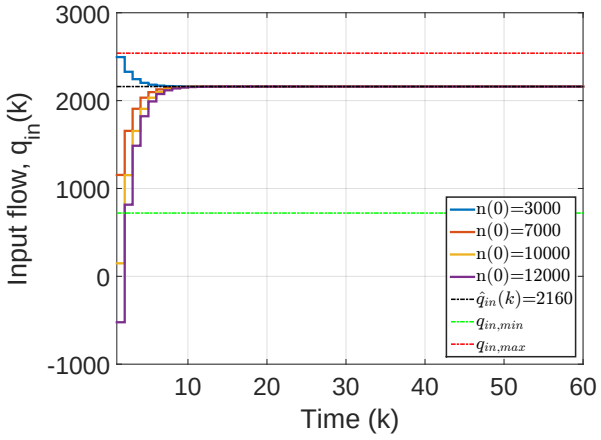


(e) Gate 5

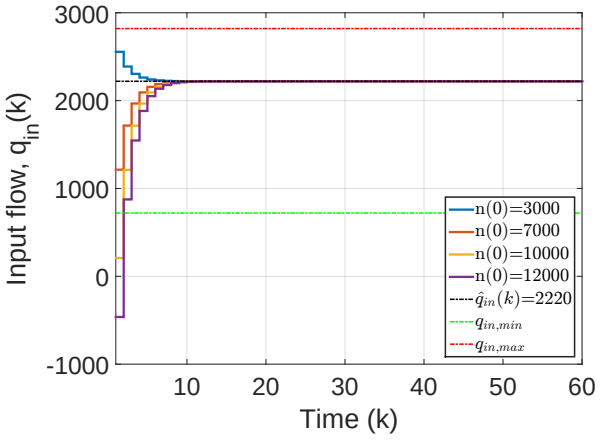


(f) Gate 6

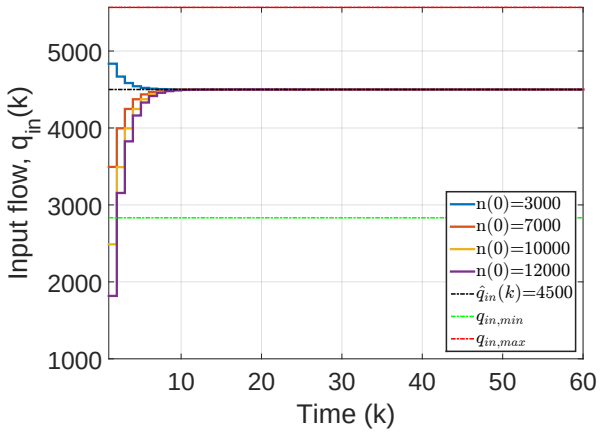
Figure 4.5: Version II: Control trajectories for gates 1–6.



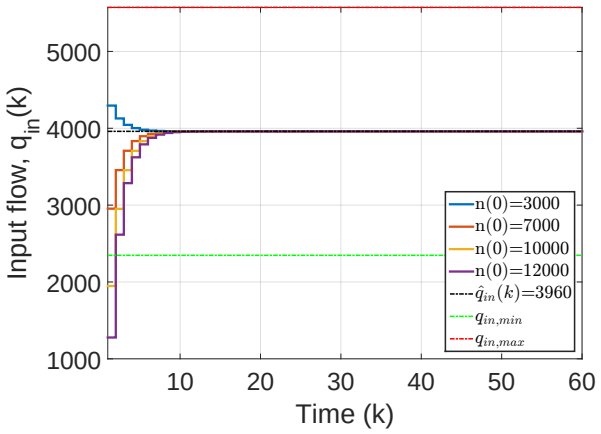
(a) Gate 7



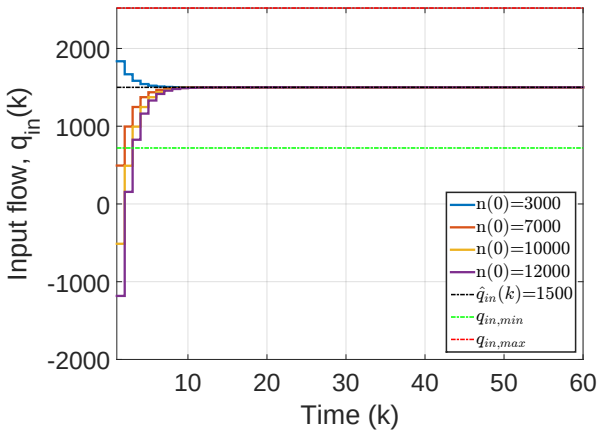
(b) Gate 8



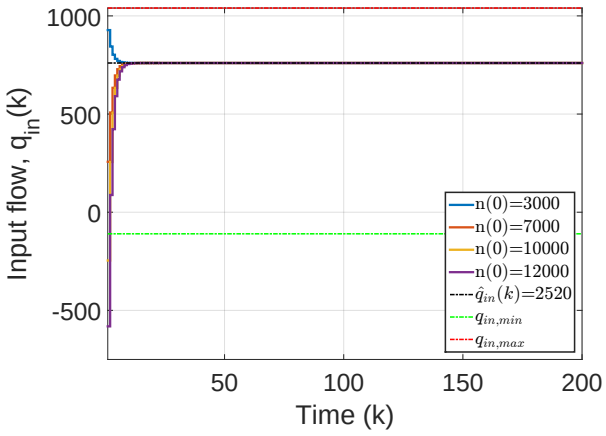
(c) Gate 9



(d) Gate 10

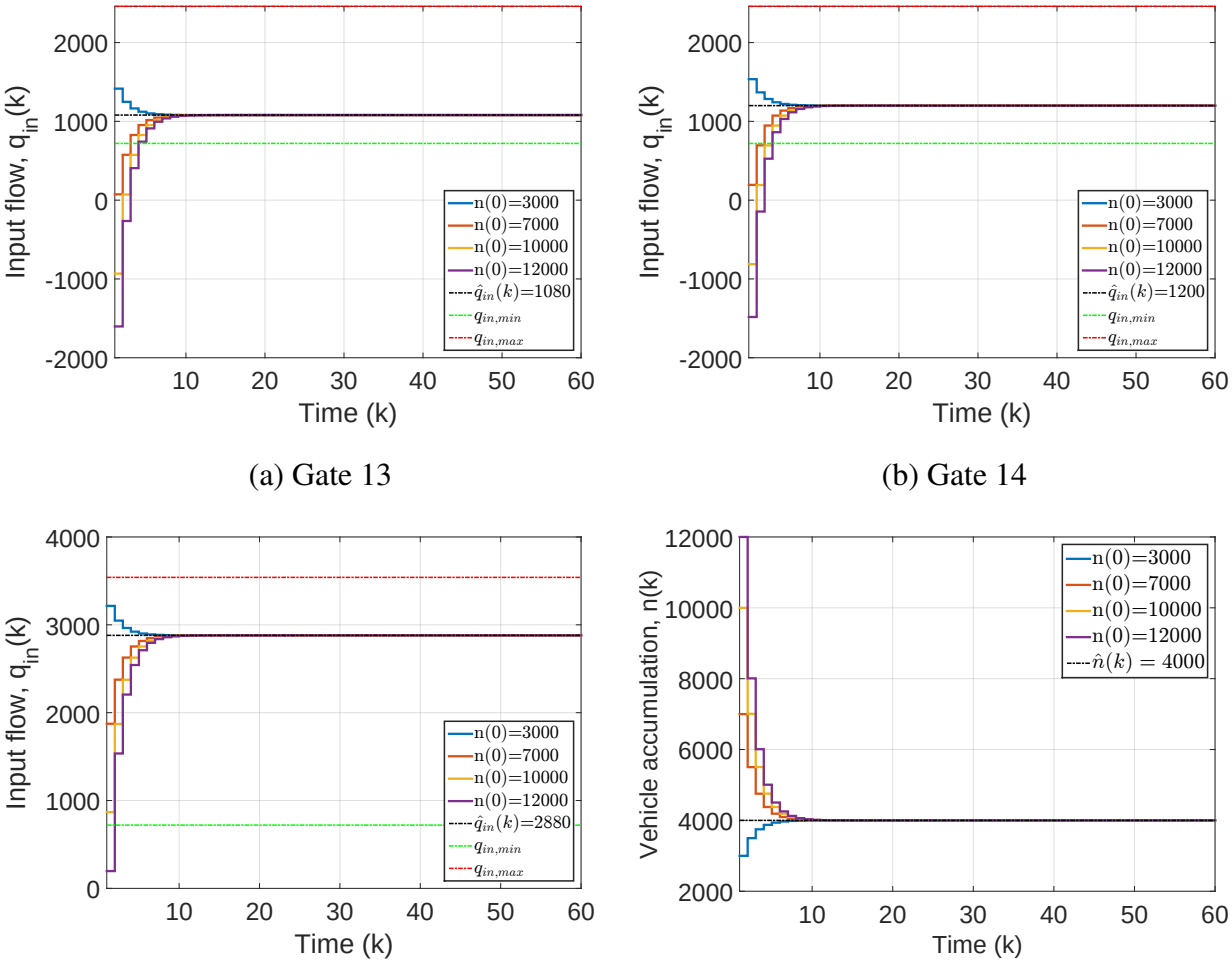


(e) Gate 11



(f) Gate 12

Figure 4.6: Version II: Control trajectories for gates 7–12.



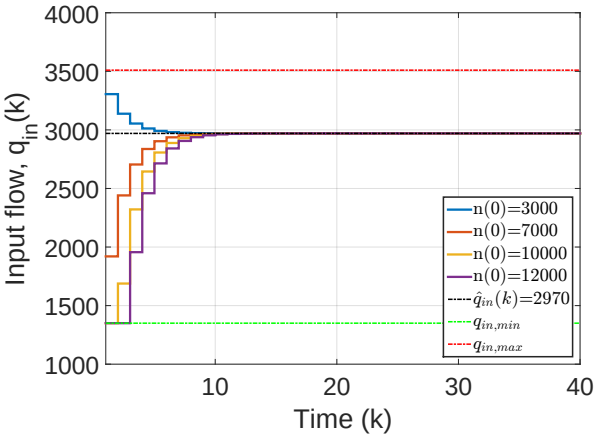
(a) Gate 13

(b) Gate 14

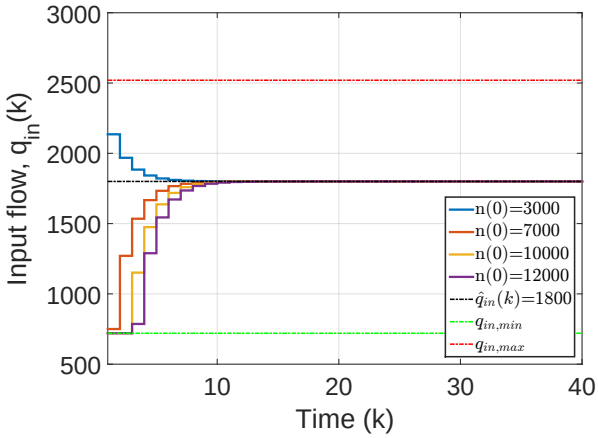
(c) Gate 15

(d) Protected Network (PN)

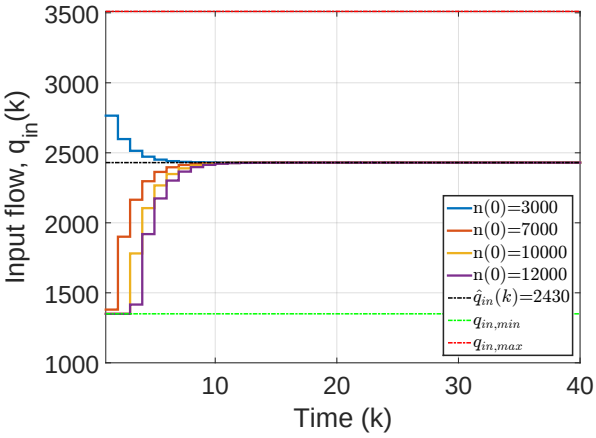
Figure 4.7: Version II: Control trajectories for gates 13–15; state trajectory of PN.



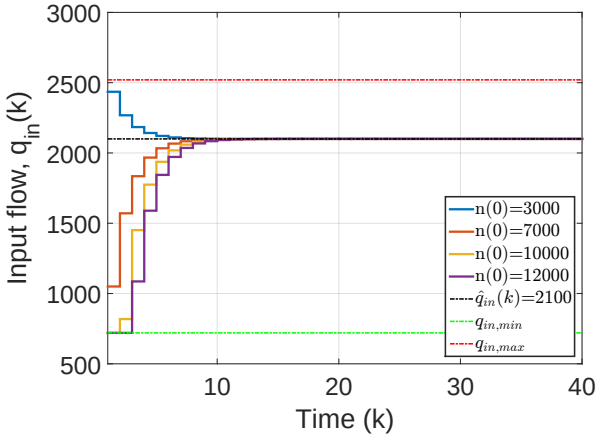
(a) Gate 1



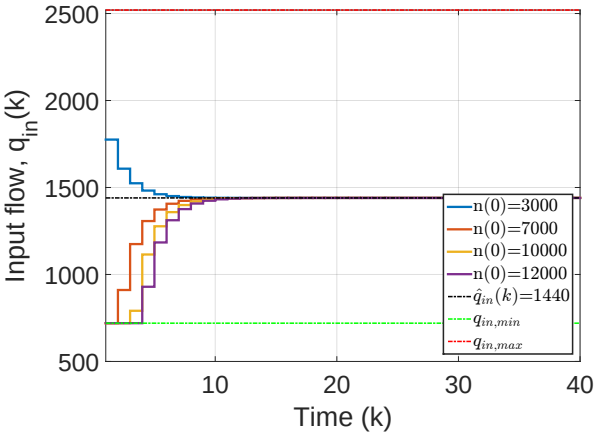
(b) Gate 2



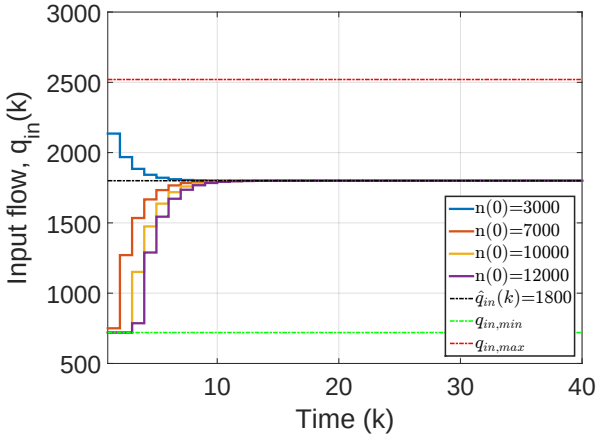
(c) Gate 3



(d) Gate 4

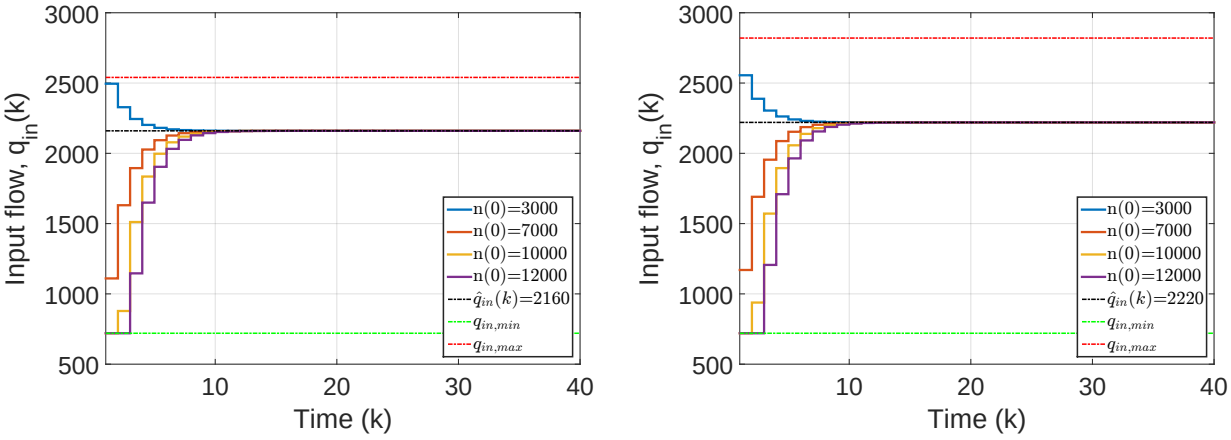


(e) Gate 5



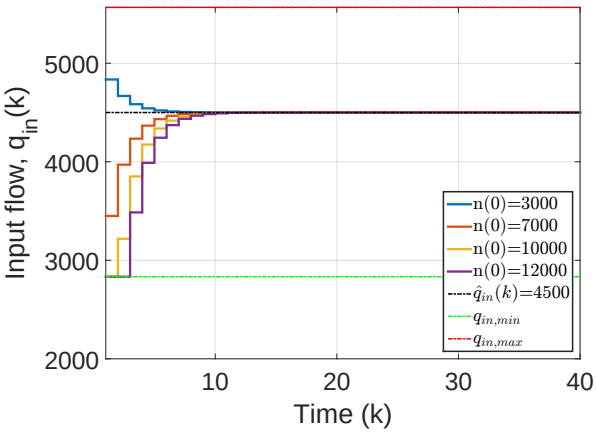
(f) Gate 6

Figure 4.8: Version II (constrained): Control trajectories for gates 1–6.

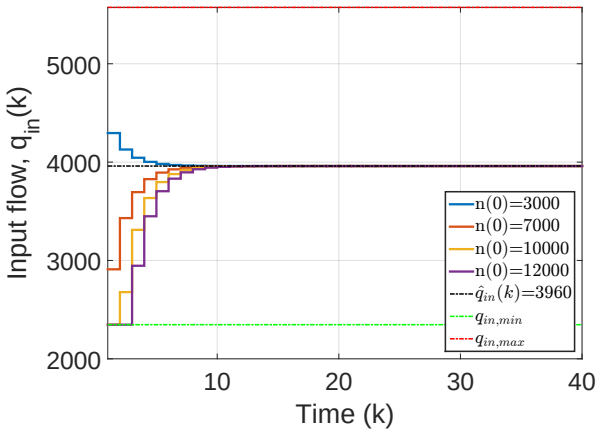


(a) Gate 7

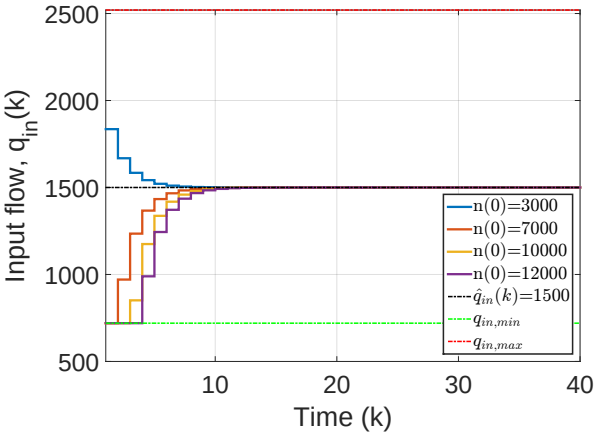
(b) Gate 8



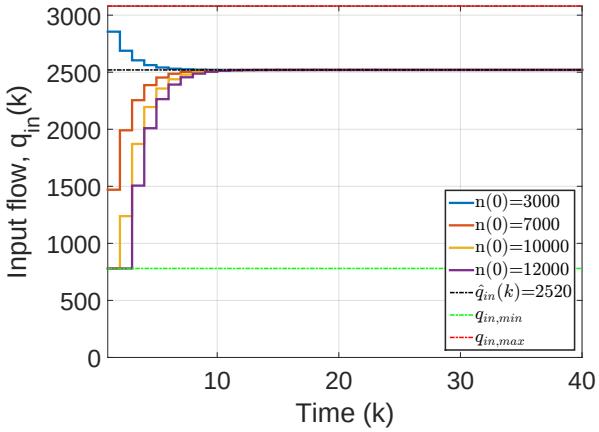
(c) Gate 9



(d) Gate 10

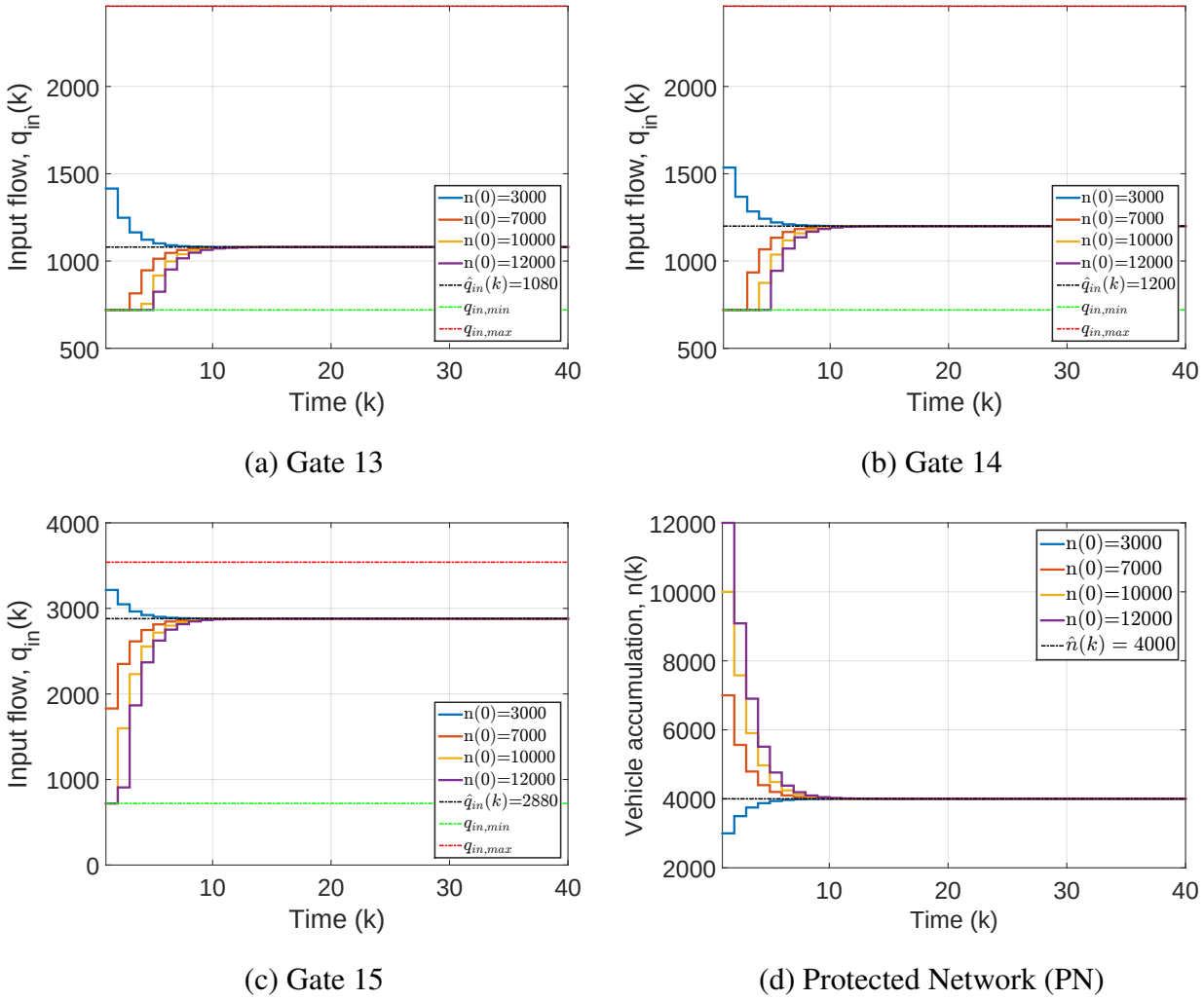


(e) Gate 11



(f) Gate 12

Figure 4.9: Version II (constrained): Control trajectories for gates 7–12.



(a) Gate 13

(b) Gate 14

(c) Gate 15

(d) Protected Network (PN)

Figure 4.10: Version II (constrained): Control trajectories for gates 13–15 and state trajectory of PN.

4.5.3.2 Multi-gated perimeter flow control: Open-loop control results

Several tests were conducted in order to investigate the behaviour of the proposed multi-gated control for different scenarios. The scenarios were created by assuming more or less high initial queues $\ell_o(0)$ in the fifteen origin links of the protected network while the protected network area operating in the congested regime, i.e. its state $n(0) > 6000$ veh. The optimisation horizon for each scenario is 2 h (40 cycles).

The calculated optimal state and control trajectories demonstrate the efficiency of the proposed multi-gated control to solve the perimeter flow control problem with queue dynamics. Figures 4.11–4.16 show some obtained trajectories for a heavy scenario with $\ell_o(0) = 0.7\ell_{o,\max}$, $\forall o = 1, \ldots, 15$ and two initial states in the congested regime of the fundamental diagram $n(0) = 7000$ veh and $n(0) = 12000$ veh (extreme case). For the initial value of vehicle accumulation $n(0)$, see the legend in each subfigure. Tests were conducted with and without

external demand flows at origin links, denoted with "s+d" and "s-d", respectively. It should be noted that despite the same level of saturation (70%) is used for all origin links, the corresponding vehicle queues observed vary due to different geometric characteristics. The main observations are summarised in the following remarks:

- The proposed strategy manages to stabilise the vehicle accumulation of the protected network around its desired state $\hat{n} = 4000$ veh for all initial states (even in the extreme case) and cases with and without disturbances (see Figure 4.11(a)).

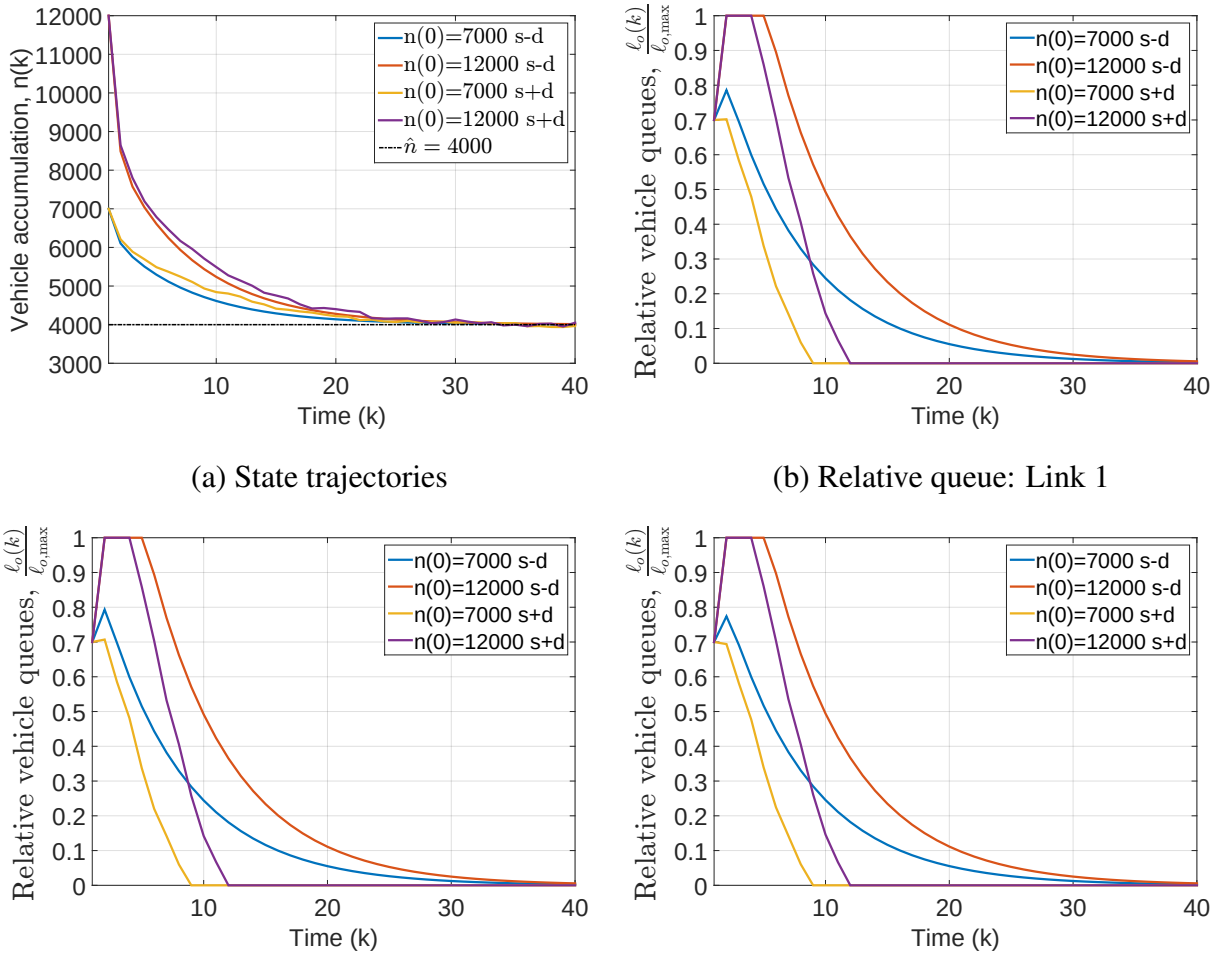
- The proposed multi-gated control strategy manages to dissolve the initial origin link queues in a balanced way (see Figures 4.11(b)–4.11(d), 4.12 and 4.13) and thus, the desired control objective of queue balancing and equity for drivers using different gates to enter the protected network area is achieved.

- The proposed strategy manages to stabilise all input flows to their desired values $\hat{\mathbf{q}}$ (corresponding to the nominal signal plan in Table 4.1) in the steady state, i.e., where $\hat{n} = 4000$ and system's throughput is maximised (see Figures 4.14–4.16, notice the different reference points $\hat{q}_o$ in each subfigure).

- The input flows ordered by the multi-gated perimeter flow control strategy have different trajectories and characteristics (see control trajectories in Figures. 4.14 – 4.16). This confirms that an equal distribution of ordered flows to corresponding junctions is not optimal, as largely assumed in previous studies. As can be seen, the proposed strategy determines optimally distributed input flows (or feasible entrance link green times) by taking into account the individual geometric characteristics of the origin links as well as minimum and maximum constraints.

- The input flows ordered by the multi-gated perimeter flow control adapting well when link dynamics are considered. For example when queues are taken into account, the trajectories at gate 1 allow more flows into protected network in order to manage the developed queues at link while still control the accumulation in protected network to be within desired value (see Figure 4.15(a) and 4.11(a)). In contrast, without information of queues at link, the perimeter flow control are inflexible and it restrict the input flow whenever $n > 4000$veh (see Figure 4.8(a)). This observation show that significant improvements can be reached through considering link dynamics and gating should not be restricted to optimisation within protected network only.

- It is evident that excessive demand and high initial queues at origin links, coupled with the applied control, causes congestion shortly after the beginning of the time horizon. At the same time the protected network is operating in the congested regime ($n(0) = 7000$ veh or $n(0) = 12000$ veh). As can be seen, the multi-gated control strategy first restricts



(a) State trajectories

(b) Relative queue: Link 1

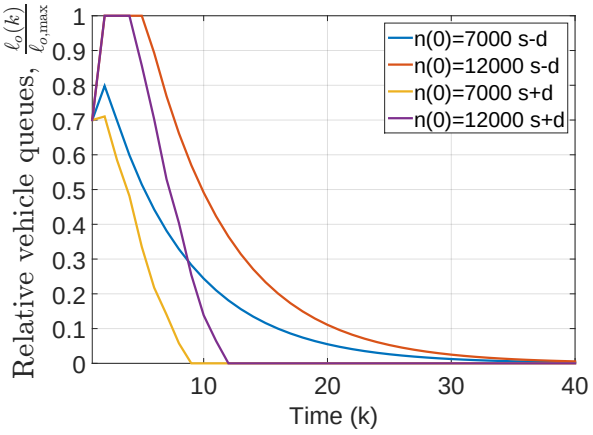
(c) Relative queue: Link 2

(d) Relative queue: Link 3

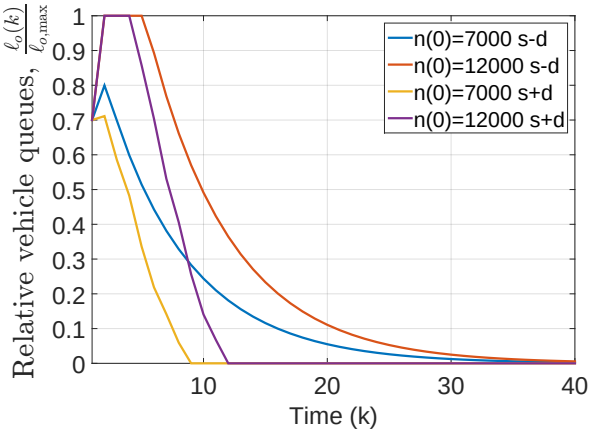
Figure 4.11: (a) State trajectories of PN; (b)–(d): Relative vehicle queues at origin links 1–3 for different initial states with and without disturbances (open-loop control results).

the high initial queues at origin links to flow into the oversaturated protected network area and then, in order to manage the developed long queues therein (in some cases reach the upper bounds), it gradually increases the input flows. Note that for some gates (7, 8 and 9) bound constraints are activated for a certain time period.

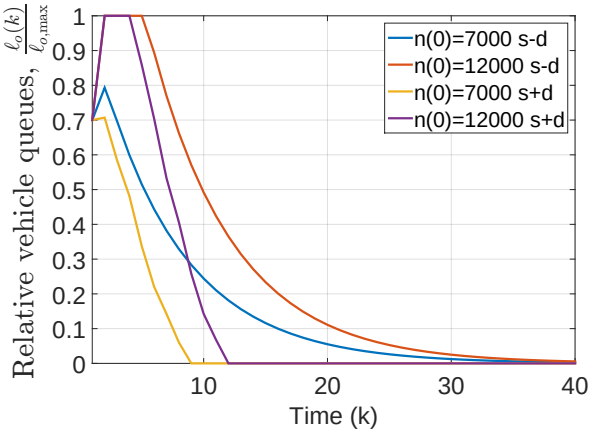
The obtained open-loop control results provided us with some good evidence on the effectiveness of the proposed multi-gated perimeter flow control strategy to solve the perimeter flow stabilisation problem with entrance link dynamics and operational constraints. The next section embeds this open-loop optimisation problem in a closed-loop responsive environment via rolling-horizon (model-based predictive control).



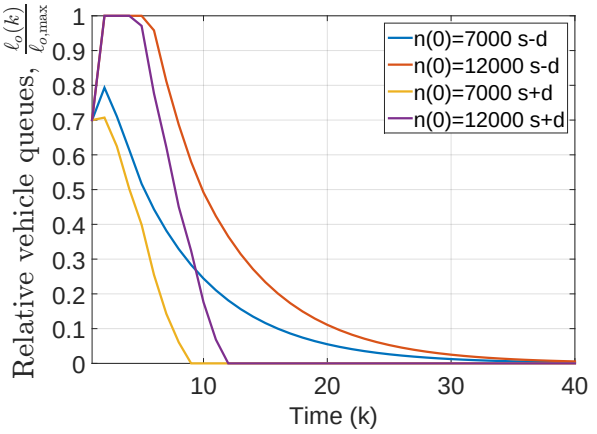
(a) Relative queue: Link 4



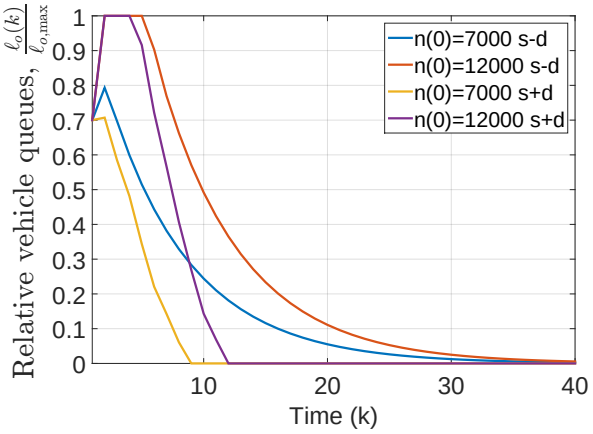
(b) Relative queue: Link 5



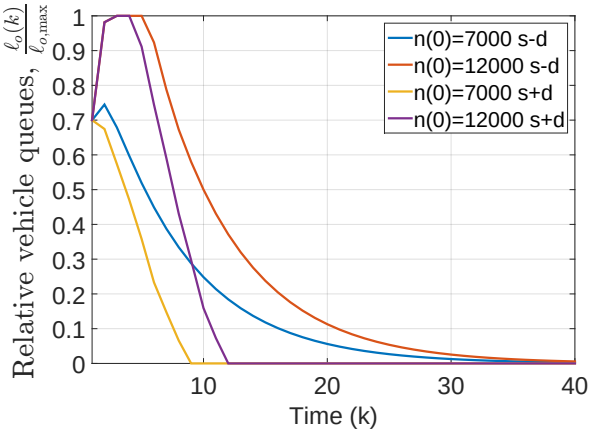
(c) Relative queue: Link 6



(d) Relative queue: Link 7

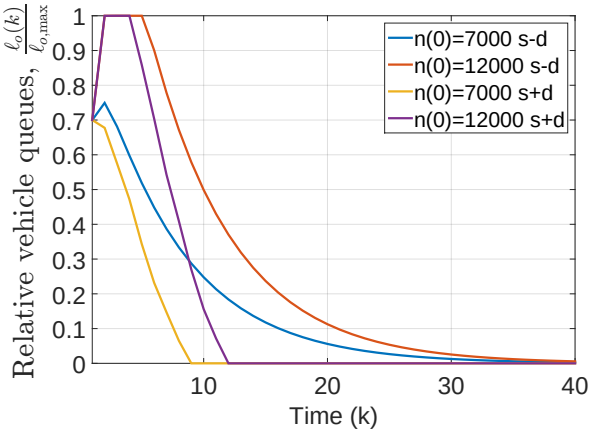


(e) Relative queue: Link 8

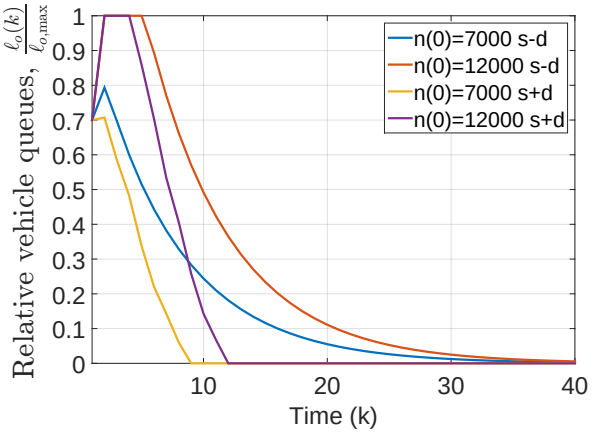


(f) Relative queue: Link 9

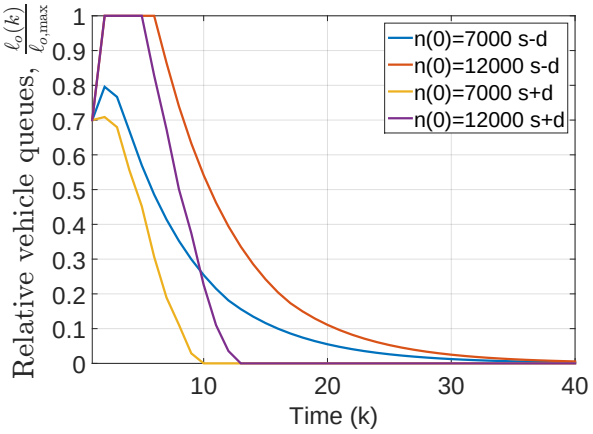
Figure 4.12: Relative vehicle queues at origin links 4–9 for different initial states with and without disturbances (open-loop control results).



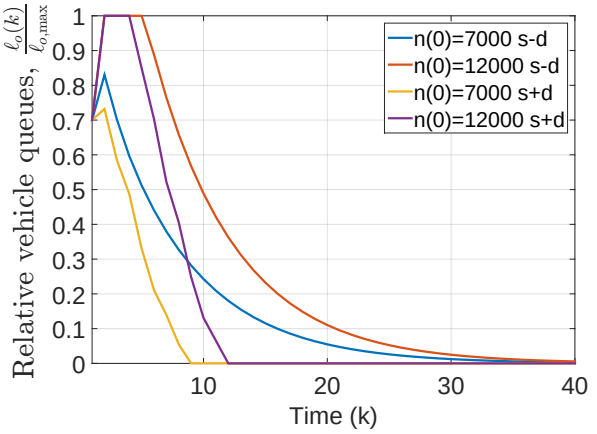
(a) Relative queue: Link 10



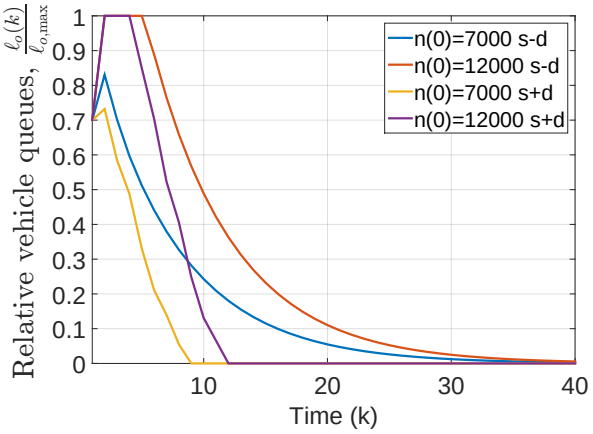
(b) Relative queue: Link 11



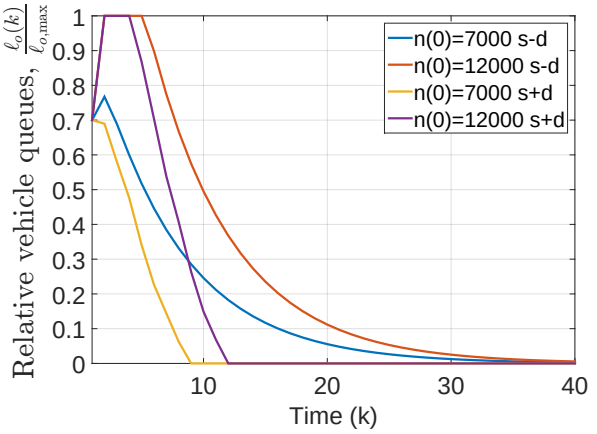
(c) Relative queue: Link 12



(d) Relative queue: Link 13

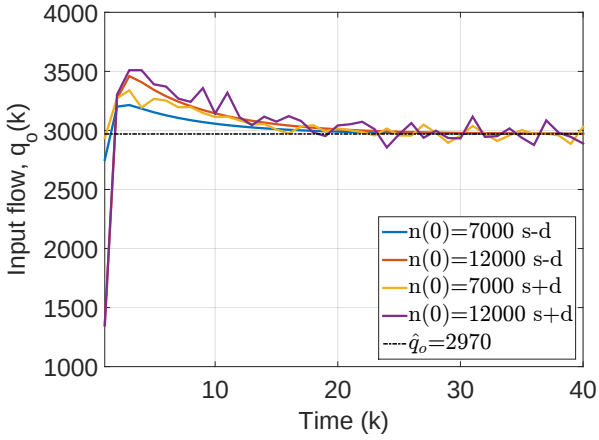


(e) Relative queue: Link 14

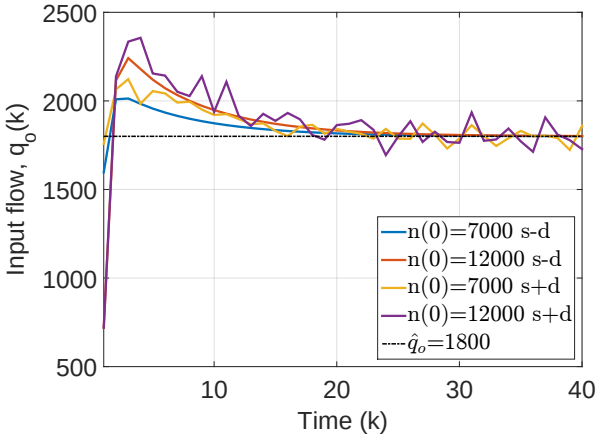


(f) Relative queue: Link 15

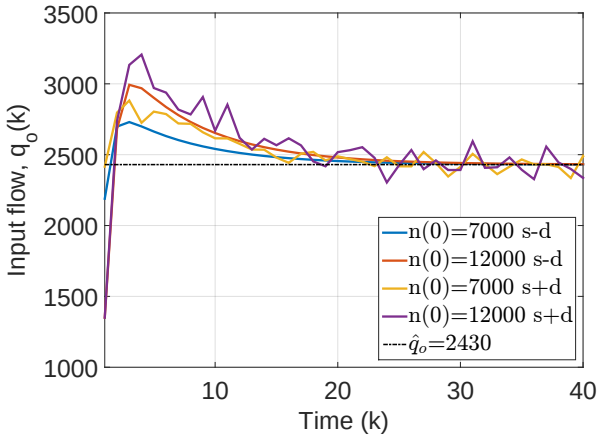
Figure 4.13: Relative vehicle queues at origin links 10–15 for different initial states with and without disturbances (open-loop control results).



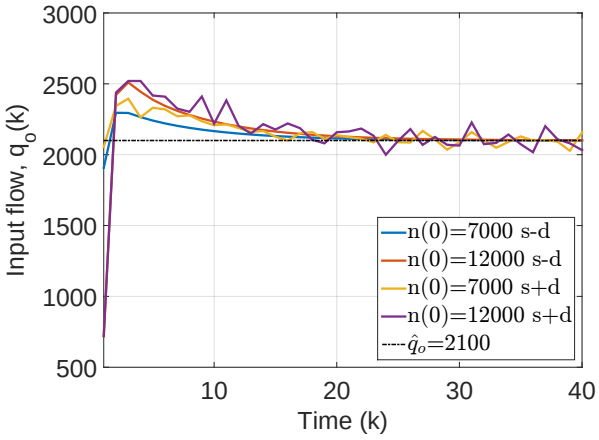
(a) Input flow: Gate 1



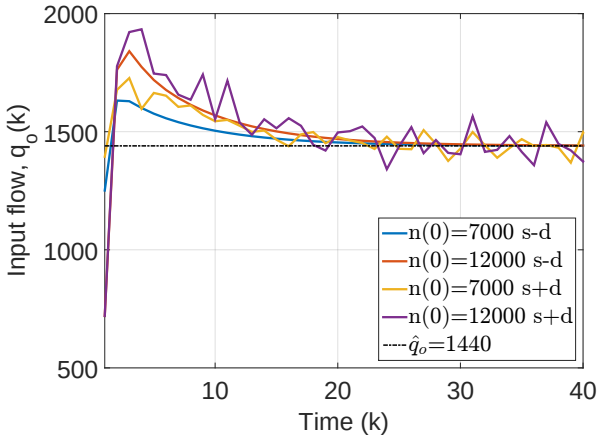
(b) Input flow: Gate 2



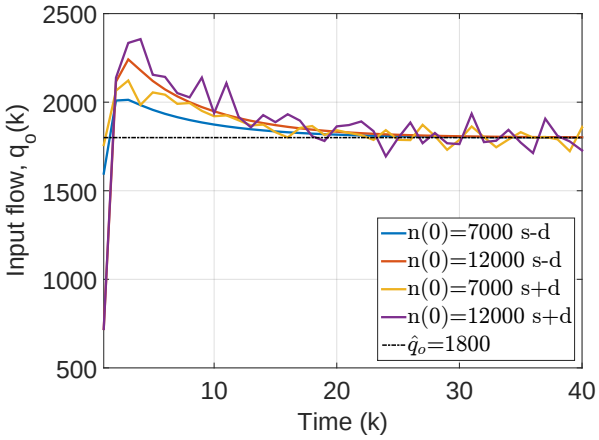
(c) Input flow: Gate 3



(d) Input flow: Gate 4

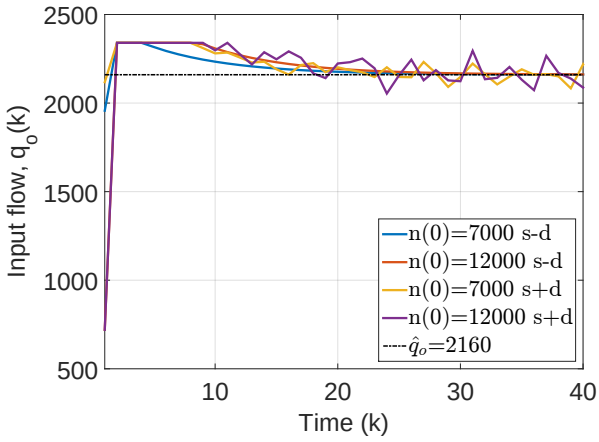


(e) Input flow: Gate 5

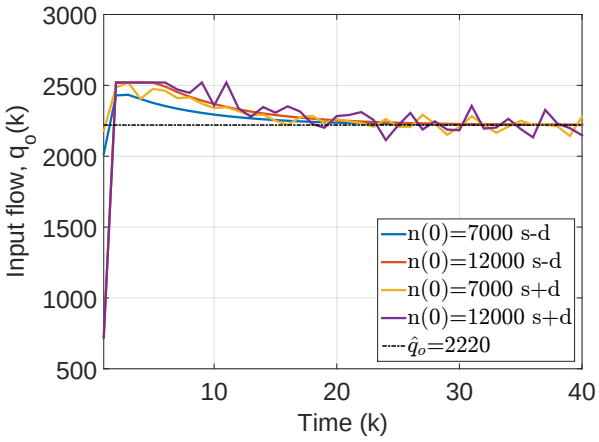


(f) Input flow: Gate 6

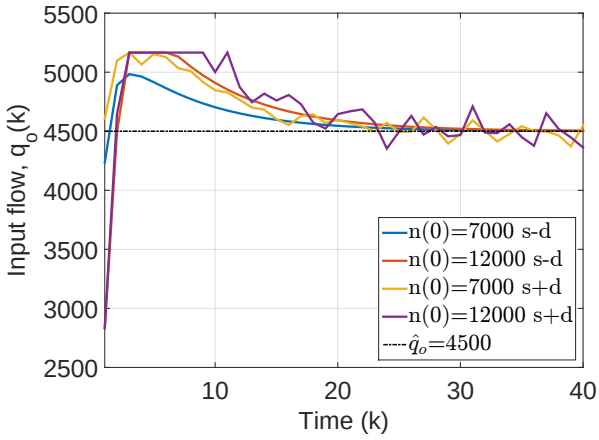
Figure 4.14: Control trajectories of MGC at gates 1–6 for different initial states with and without disturbances (open-loop control results).



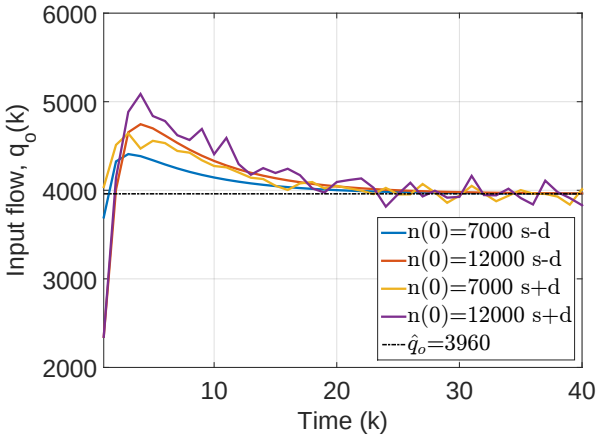
(a) Input flow: Gate 7



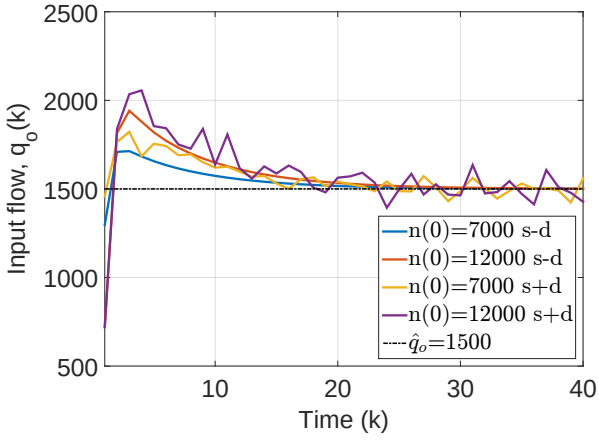
(b) Input flow: Gate 8



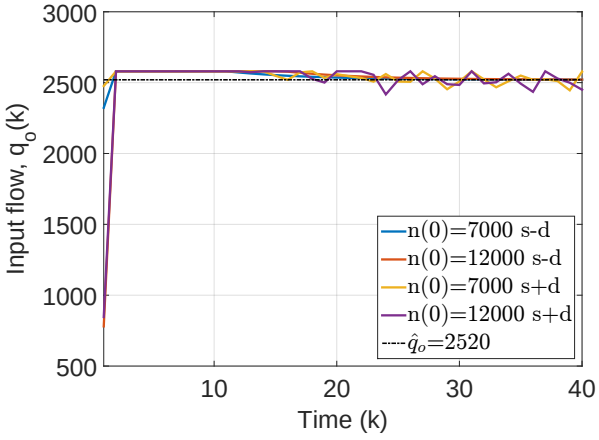
(c) Input flow: Gate 9



(d) Input flow: Gate 10

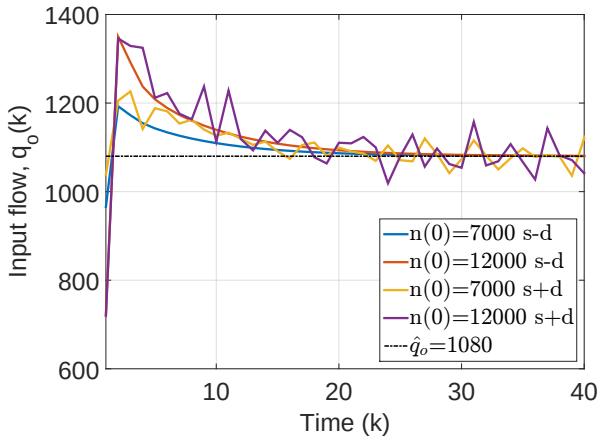


(e) Input flow: Gate 11

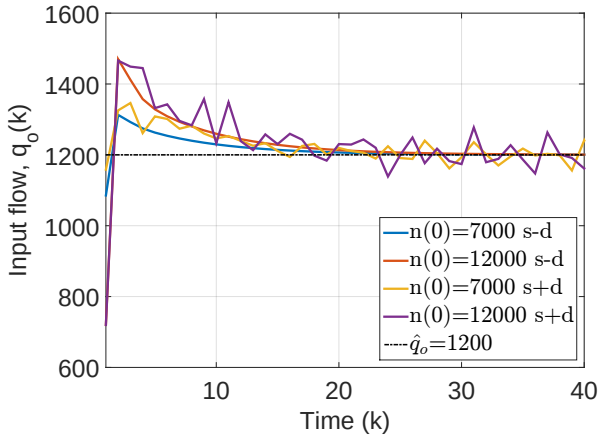


(f) Input flow: Gate 12

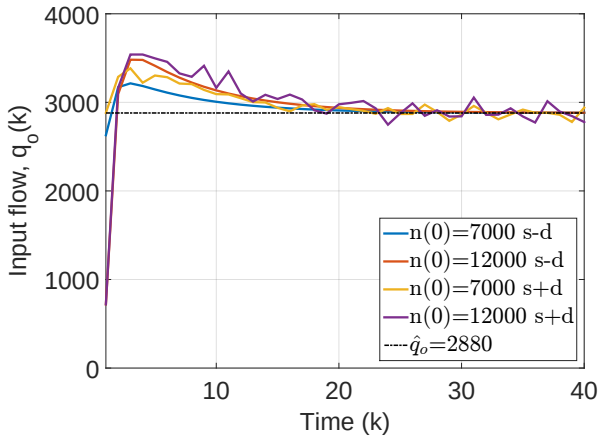
Figure 4.15: Control trajectories of MGC at gates 7–12 for different initial states with and without disturbances (open-loop control results).



(a) Input flow: Gate 13



(b) Input flow: Gate 14



(c) Input flow: Gate 15

Figure 4.16: Control trajectories of MGC at gates 13–15 for different initial states with and without disturbances (open-loop control results).

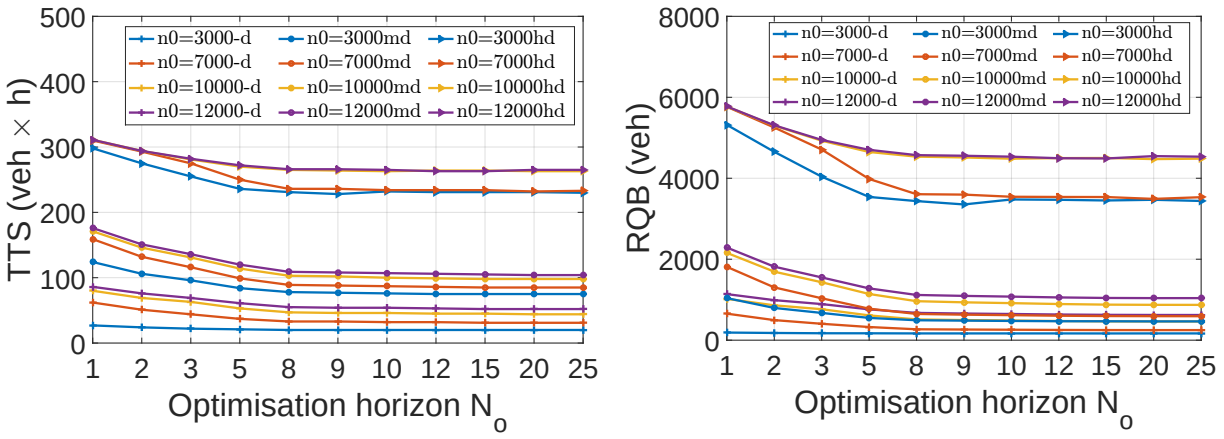
4.5.3.3 Multi-gated perimeter flow control: Closed-loop control results

This section presents the proposed multi-gated perimeter flow control embedded in a closed-loop structure via rolling-horizon (model predictive control). Model-based control is a repetitive optimisation scheme, where at each time step an open-loop constrained optimisation problem with finite horizon N_o and predicted demands $\mathbf{d}(k)$ over a prediction horizon N_p is solved, then only the first control move is applied to the plant and the procedure is carried out again. Here $N_o = N_p$ is assumed but an appropriate value of N_o should be specified. In principle a short optimisation horizon would lead to myopic decisions, while a long one demands for high computational effort, provided that the number of decision variables is increased by $N_o \times$ (# controls + # states). Then the MGC strategy is applied for two demand scenarios (medium and high) to the protected network area of downtown San Francisco.

Specification of the optimisation horizon N_o

Twelve scenarios were defined in order to investigate the behaviour of the multi-gated perimeter flow control strategy under different initial states and demand scenarios. The twelve scenarios composed of four initial states in the uncongested and congested regimes (near gridlock traffic conditions) of the fundamental diagram $n(0) = \{3000, 7000, 10000, 12000\}$ veh and three different demand scenarios namely no external demand ($\mathbf{d} = \mathbf{0}$), medium demand, and high demand. For the medium and high scenarios trapezoidal demands have been used for $d_o(k)$, $o = 1, \ldots, 15$, $k = 0, \ldots, N_p - 1$ over the predicted horizon of $N_o = N_p$. To capture the uncertainty of the (scaled) fundamental diagram, particularly when the network is operating in the congested regime (notice the noise for $n > 6000$ veh), d_n is selected to vary gradually with respect to $n(k)$ in the range $[-5000, 5000]$ veh/h for $n > 6000$ veh. The rolling-horizon strategy is run with different optimisation horizons $N_o = \{1, 2, 3, 5, 8, 9, 10, 12, 15, 20, 25\}$ in order to investigate the impact of N_o on the control performance. For each of the twelve scenarios and for each demand scenario, two evaluation criteria are calculated for comparison, TTS and RQB as in (4.20) and (4.21), respectively.

Figure 4.17 displays the obtained TTS and RQB results for the rolling-horizon multi-gated perimeter traffic flow control approach for different optimisation horizons $N_o = \{1, 2, 3, 5, 8, 9, 10, 12, 15, 20, 25\}$. It can be seen that for $N_o \geq 10$ there are no significant deviations of the evaluation criteria for different optimisation horizons N_o even for the high-demand scenarios. Tables 4.3 and 4.4 provide additional evidence on the evolution of the considered assessment criteria. The assessment criteria at the gated links are seen to improve as N_o increased in some scenarios. In particular, for the most congested accumulation ($n(0) = 12000$ veh) and with high demand, the most satisfactory results with respect to both evaluation criteria are obtained with $N_o = 15$ (equivalent to 0.75 h = 45 minutes). Hence $N_o = 15$ equivalent to 45 minutes is a reasonable choice. In principle, a satisfactory optimisation horizon is the one that is in the order of the time needed to travel through the network as suggested in [3].



(a) TTS (b) RQB

Figure 4.17: Sensitivity analysis and performance of MGC for different optimisation horizon N_o and demand scenarios: (a) TTS within the protected network and outside network area; (b) RQB within the protected network and outside network area

Simulation results for medium demand

A medium demand scenario is first investigated with the following characteristics: (a) To capture the uncertainty of the NFD (particularly in the congested regime) a random demand is applied to the capacity flow in the range $[-5000, 5000]$ veh/h; (b) The demand applied at links is uniformly distributed for $k \leq 40$ and zero otherwise, i.e. for $41 \leq k \leq 60$. Figures 4.18–4.19 illustrate the pattern of the demand used at the protected network (NFD) and origin links (gates).

Several tests were conducted under different scenarios in order to investigate the behaviour of the proposed MGC scheme. First, scenarios were created with random (but high) initial queues $\ell_o(0)$ in the fifteen origin links of the protected network. The vector of initial queues at origin links is $\ell_o(0) = [0.7; 0.7; 0.8; 0.8; 0.45; 0.5; 0.7; 0.8; 0.9; 0.85; 0.8; 0.8; 0.9; 0.7; 0.8]$. Second, scenarios were created for the protected network area to operate in four different traffic conditions: (a) uncongested; (b) semi-congested; (c): congested; and, (d) oversaturated (near gridlock). The different traffic conditions represented by different initial accumulations $n(0)$ ranging in $[3000, 12000]$ veh. The simulation horizon for each scenario is 3 h (60 cycles).

Figures 4.23–4.25 and 4.20–4.22 show the obtained control trajectories and relative vehicle queues (states) for four initial states of the fundamental diagram $n(0) = [3000, 7000, 10000, 12000]$ veh. In all simulations the initial queues $\ell_o(0)$ is assumed as above. Note that y-axis in Figures 4.20–4.22 indicate the minimum and maximum allowable values of each q_o. Tests were conducted with and without external demand flows at origin links, denoted with "+d" and "-d", respectively.

For the scenario without external demand flows, MGC manages to stabilise around its desired accumulation at $k \approx 30$ (Figure 4.25(d) dashed lines), while in the constrained control without entrance link dynamics (Version II) at $k \approx 10$. This difference is attributed to the time MGC

Table 4.3: Total time spent (TTS) within the protected network area.

N_o	n(0) = 3000			n(0) = 7000			n(0) = 10000			n(0) =12000		
	no demand	medium demand	high demand	no demand	medium demand	high demand	no demand	medium demand	high demand	no demand	medium demand	high demand
	TTS	TTS	TTS	TTS	TTS	TTS	TTS	TTS	TTS	TTS	TTS	TTS
1	12116	12938	14955	12924	13744	15536	13462	14225	15889	13789	14529	16156
2	12116	12939	15234	12925	13747	15947	13488	14260	16293	13842	14589	16575
3	12116	12939	15293	12925	13748	16041	13523	14303	16424	13871	14625	16692
5	12116	12939	15303	12925	13748	16097	13531	14337	16503	13911	14684	16793
8	12116	12939	15306	12925	13748	16110	13531	14350	16535	13926	14714	16834
9	12116	12939	15307	12925	13748	16108	13531	14352	16542	13928	14721	16842
10	12116	12939	15297	12925	13748	16112	13531	14352	16550	13929	14723	16848
12	12116	12939	15293	12925	13748	16105	13531	14353	16549	13931	14727	16860
15	12116	12939	15293	12925	13748	16103	13531	14354	16550	13932	14730	16864
20	12116	12939	15292	12925	13748	16108	13531	14354	16556	13936	14734	16857
25	12116	12939	15296	12925	13748	16104	13531	14354	16554	13936	14735	16860

Table 4.4: Total time spent (TTS) and relative queue balance (RQB) at gated links.

N_o	n(0)= 3000						n(0)=7000						n(0)=10000						n(0)=12000					
	no demand		medium demand		high demand		no demand		medium demand		high demand		no demand		medium demand		high demand		no demand		medium demand		high demand	
	TTS	RQB	TTS	RQB	TTS	RQB	TTS	RQB	TTS	RQB	TTS	RQB	TTS	RQB	TTS	RQB	TTS	RQB	TTS	RQB	TTS	RQB	TTS	RQB
1	27	188	124	1039	298	5319	62	655	159	1812	310	5766	80	1021	171	2157	311	5770	86	1138	176	2288	311	5774
2	24	178	106	796	275	4653	51	496	132	1304	293	5253	69	861	146	1692	294	5300	76	989	151	1816	294	5308
3	22	171	96	673	255	4039	44	405	116	1035	275	4709	63	765	131	1431	281	4927	69	892	136	1550	282	4947
5	21	170	84	550	236	3538	37	325	99	778	250	3982	53	612	114	1141	270	4649	61	761	120	1279	272	4702
8	20	167	78	492	231	3436	33	269	89	648	236	3606	47	513	103	963	265	4534	55	674	109	1117	266	4571
9	20	167	77	483	228	3353	33	263	88	630	236	3596	46	497	102	935	264	4514	54	658	108	1094	266	4556
10	20	167	76	475	232	3476	32	258	87	618	234	3543	46	486	100	915	263	4483	54	647	107	1076	265	4533
12	20	167	75	468	231	3468	32	253	86	604	234	3538	45	473	99	893	264	4498	53	634	106	1056	263	4495
15	20	166	75	467	231	3451	31	250	85	596	234	3536	45	465	98	878	264	4501	52	626	105	1043	263	4489
20	20	167	75	464	231	3468	31	249	85	593	232	3490	44	462	98	873	263	4473	52	623	104	1039	265	4547
25	20	167	75	464	230	3439	31	249	85	592	233	3532	44	462	98	872	263	4484	52	622	104	1039	265	4533

needs to accommodate the queues at origin links. Queues at origin links (see dashed lines in Figures 4.23–4.25) dissolve in a balanced way and thus, the control objective of queue balancing and equity for drivers using different gates to enter the protected network area is achieved.

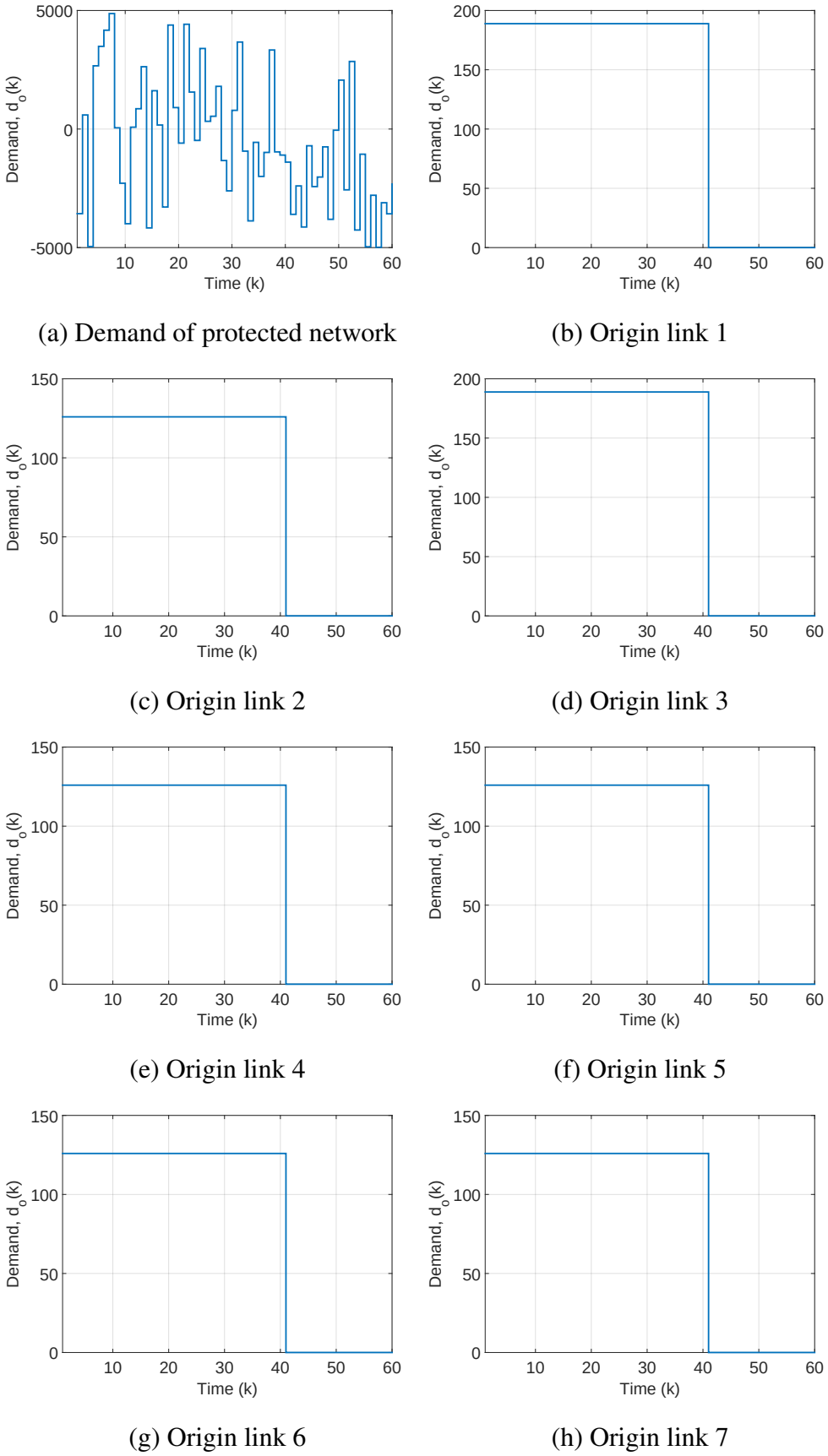
MGC is also seen to perform excellently in the case of persistent disturbances (see solid lines in Figures 4.23–4.25), though it needs more time to stabilise the system (protected and outside network areas). It should be noted that the vehicle queues observed are varying due to different geometric characteristics.

Except of queue balancing, MGC also optimally distribute the input flows to the gates and stabilise all flows to their desired $\hat{q_o}$ independent of the magnitude of disturbances. In the case of no external disturbances, the settling time is around $k = 30$ (se dashed lines in Figures 4.20–4.22) while with medium disturbances the settling time is longer around $k \approx 47$ (see solid lines in Figures 4.20–4.22). MGC respects the lower and upper bounds of input flow at each gate. The strategy allows vehicles to enter the protected network area for uncongested traffic ($n(0) < 6000$ veh) while input flow is restricted for congested conditions ($n(0) > 6000$ veh).

Simulation results for high demand

A high demand (60% higher than the medium demand studied in previous section) is introduced to further investigate the performance of the proposed MGC policy. This scenario has the following characteristics: (a) To capture the uncertainty of the NFD (particularly in the congested regime) a random demand is applied to the capacity flow in the range $[-5000, 5000]$ veh/h; (b) The demand (60% higher compared to medium) applied at links is uniformly distributed for $k \leq 40$ and zero otherwise, i.e. for $41 \leq k \leq 60$. Figures 4.26–4.27 illustrate the pattern of the demand used at the protected network (NFD) and origin links (gates). Several tests were conducted under different scenarios in order to investigate the behaviour of the proposed MGC scheme as in the previous section for comparable results.

Figures 4.28–4.30 and 4.31–4.33 show the obtained control trajectories and relative vehicle queues (states) for four initial states of the fundamental diagram $n(0) = [3000, 7000, 10000, 12000]$ veh. In all simulations the initial queues $\ell_o(0)$ is assumed as above. With high demand in the periphery of the protected network area, MGC is seen to accomplish queue balancing at all links though taking longer time to dissolve the queues compared to scenarios without demand information or with medium demand. Approximately queues at links dissolve at $k = 57$. Note that for some links (e.g. 12–14), upper bound constraints are activated for a certain time period to manage the developed long queues (see solid lines in Figures 4.29(f) and 4.30(a–b)). In addition, the proposed strategy is seen to stabilise all input flows to their desired $\hat{\mathbf{q}}$ and optimally distribute flows to the respective links. With high demand the settling time is longer approximately at $k = 57$ (see Figures 4.31–4.33). Concluding, despite the heavy incoming demand, the proposed strategy served every link in a balanced way which highlights its ability to provide equity for drivers using different gates to enter a protected network area.



(a) Demand of protected network

(b) Origin link 1

(c) Origin link 2

(d) Origin link 3

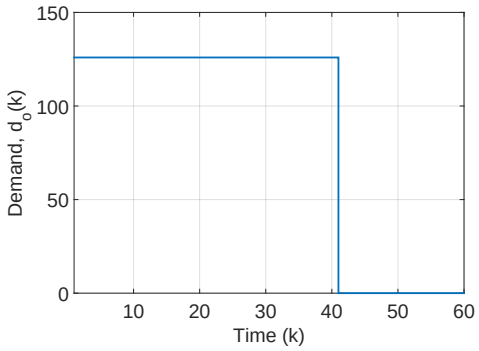
(e) Origin link 4

(f) Origin link 5

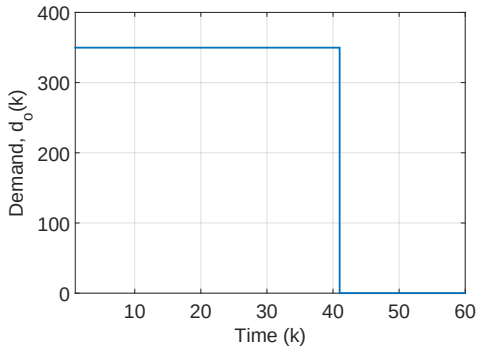
(g) Origin link 6

(h) Origin link 7

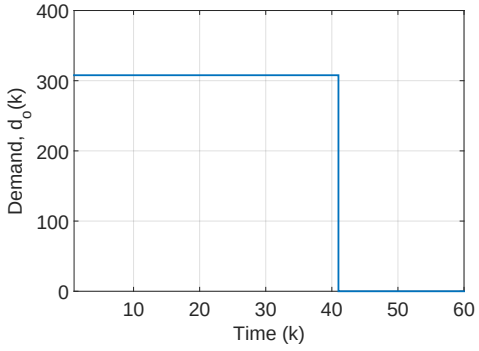
Figure 4.18: Medium demand senario; (a): Demand at PN; (b)–(f): Demand at origin links 1–7.



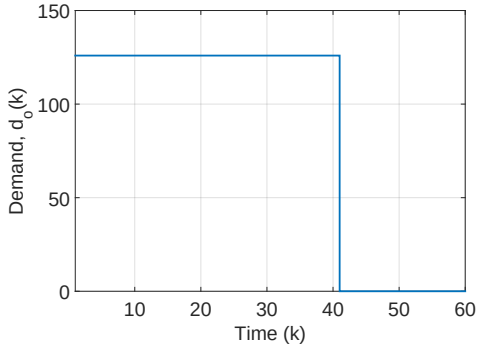
(a) Origin link 8



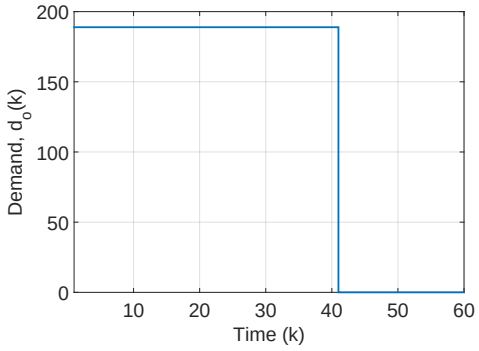
(b) Origin link 9



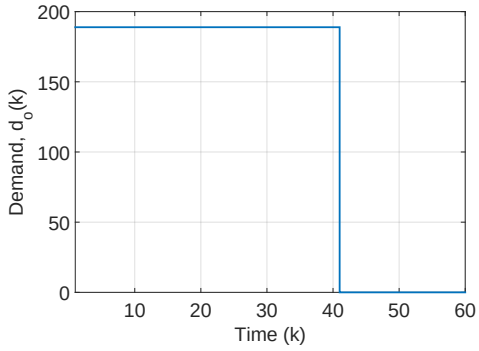
(c) Origin link 10



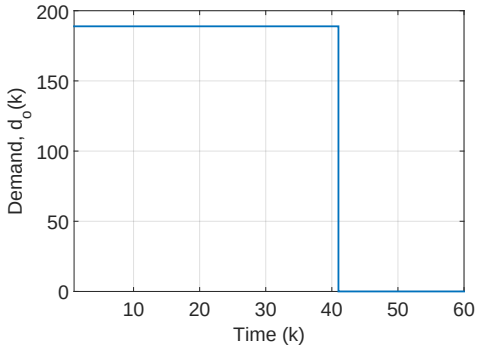
(d) Origin link 11



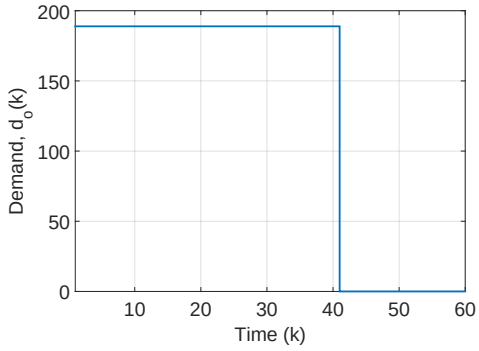
(e) Origin link 12



(f) Origin link 13

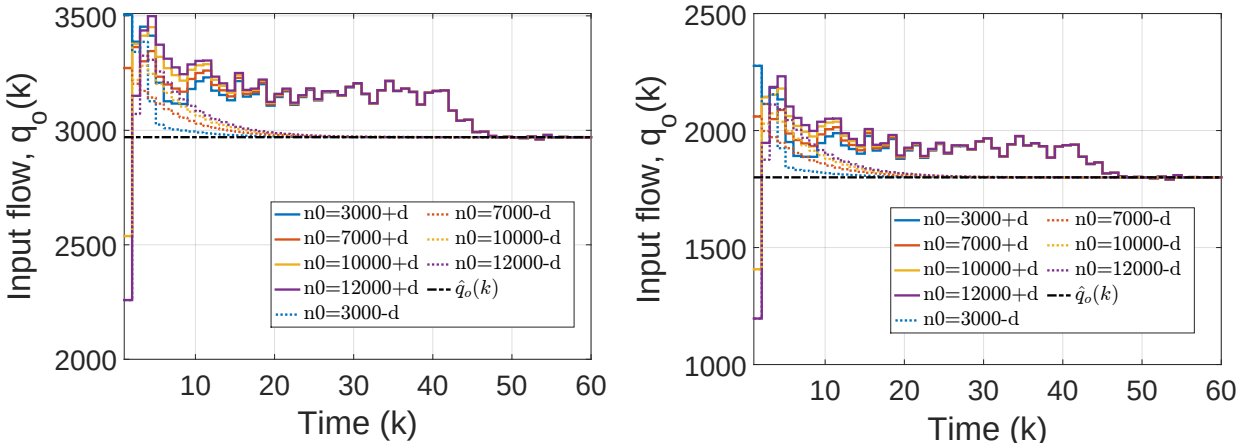


(g) Origin link 14

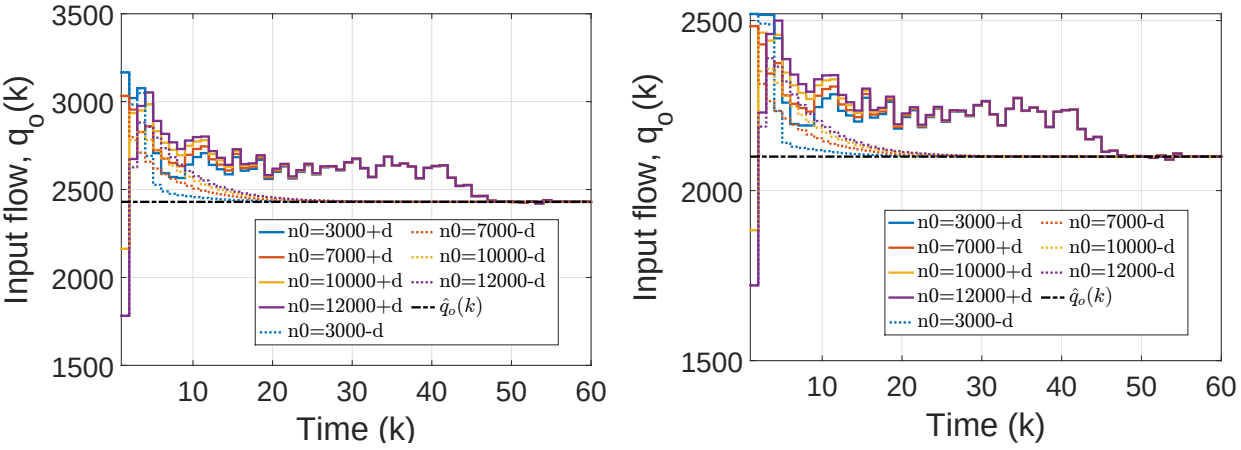


(h) Origin link 15

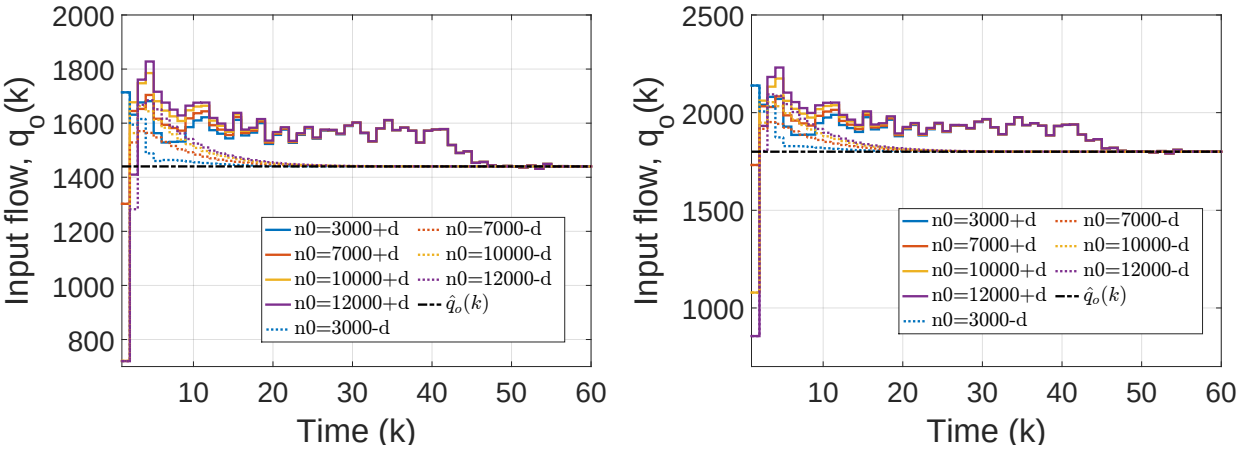
Figure 4.19: Medium demand scenario: Demand at origin links 8–15.



(a) Gate 1 (b) Gate 2

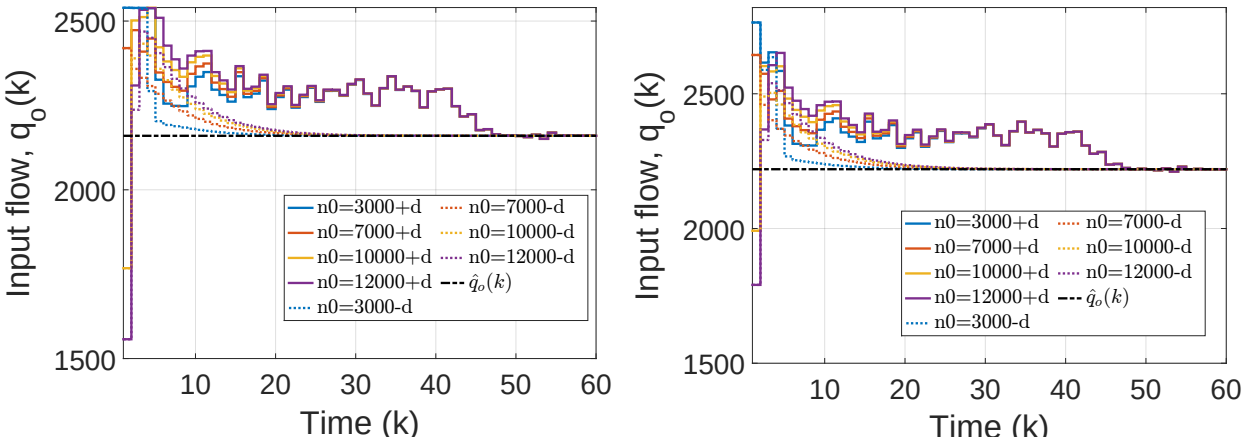


(c) Gate 3 (d) Gate 4

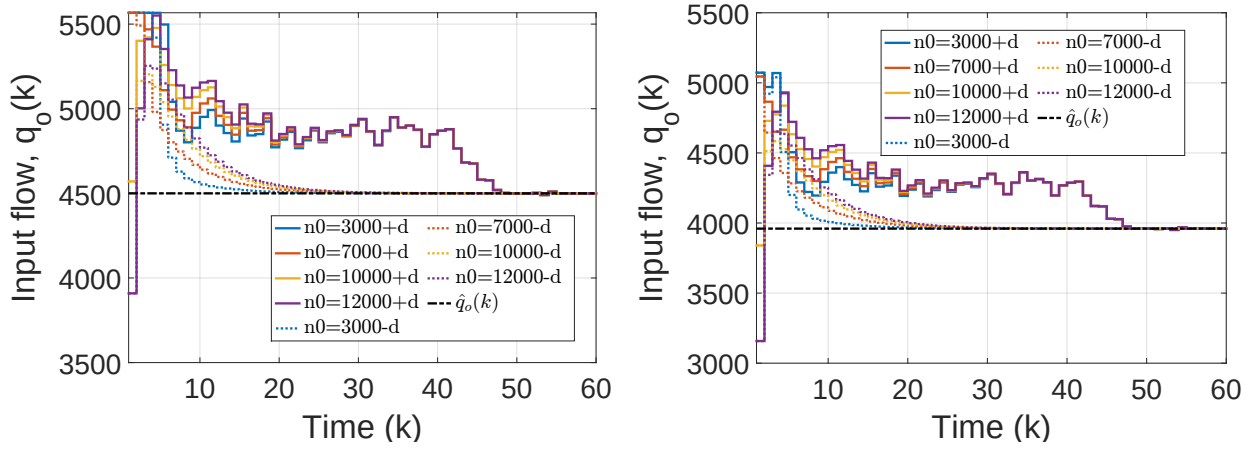


(e) Gate 5 (f) Gate 6

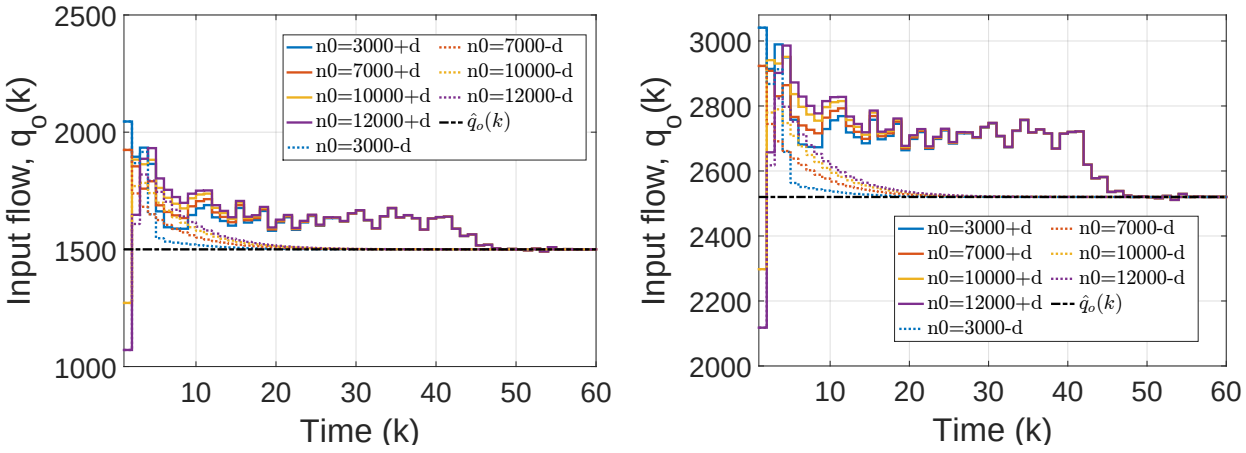
Figure 4.20: Medium demand scenario: Ordered flow for gates 1–6.



(a) Gate 7 (b) Gate 8

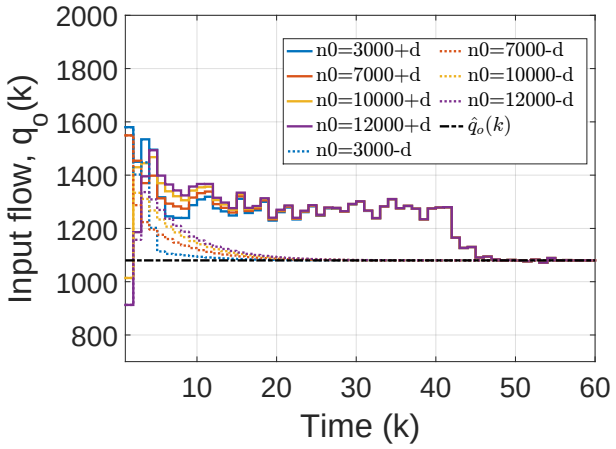


(c) Gate 9 (d) Gate 10

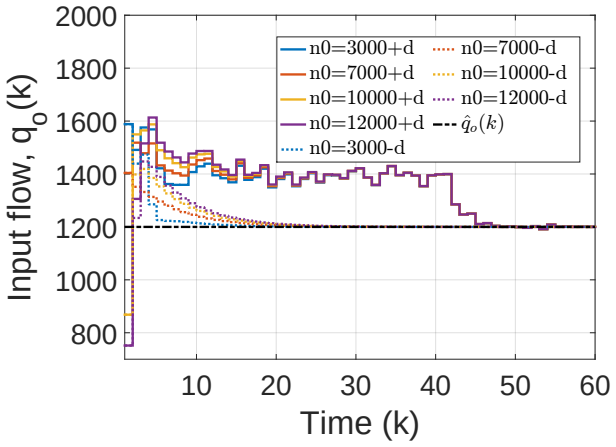


(e) Gate 11 (f) Gate 12

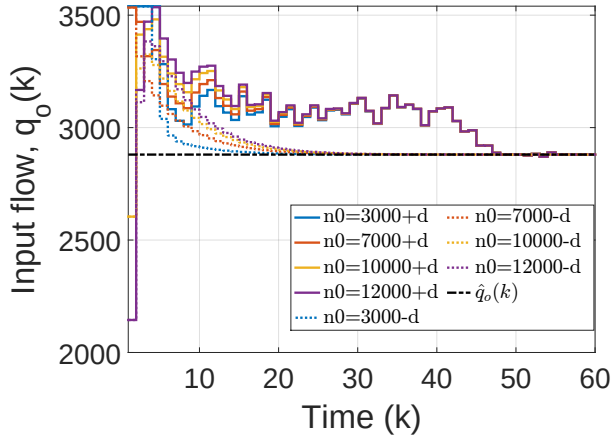
Figure 4.21: Medium demand scenario: Ordered flow for gates 7–12.



(a) Gate 13

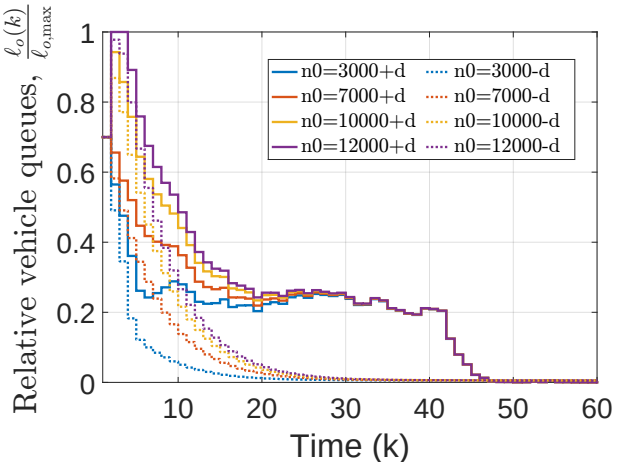


(b) Gate 14

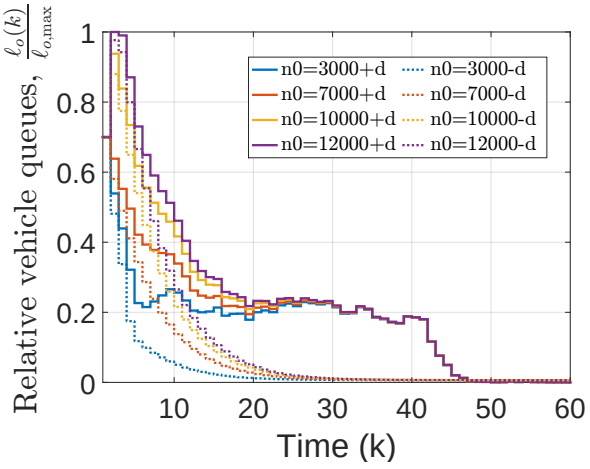


(c) Gate 15

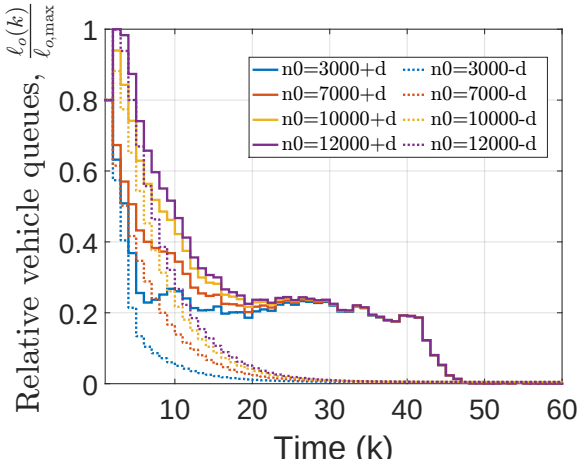
Figure 4.22: Medium demand scenario: Ordered flow for gates 13–15.



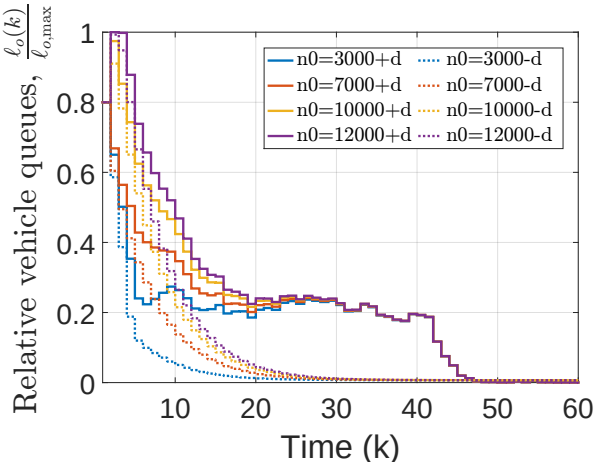
(a) Origin link 1



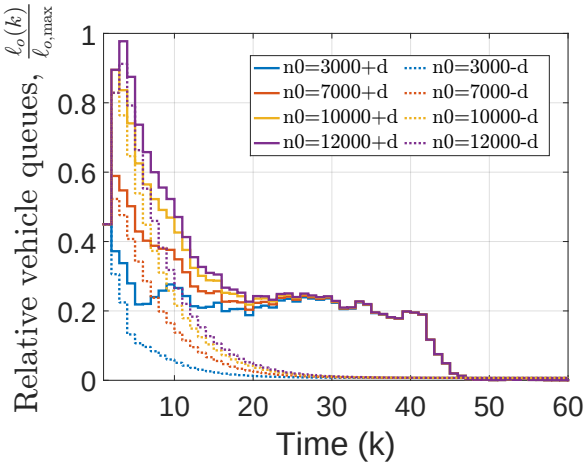
(b) Origin link 2



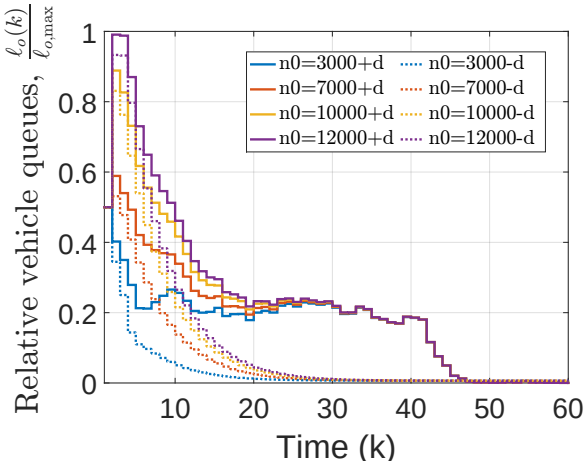
(c) Origin link 3



(d) Origin link 4

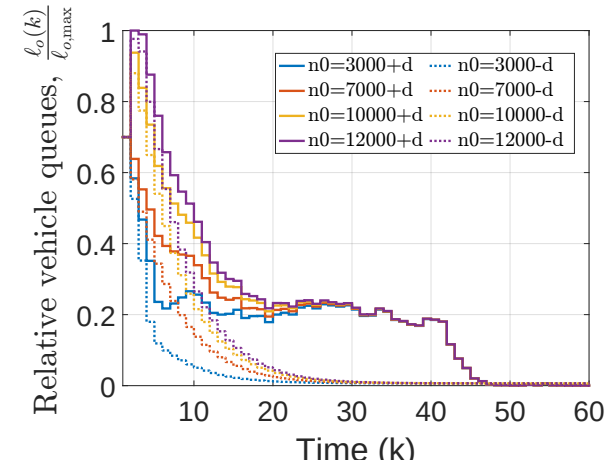


(e) Origin link 5

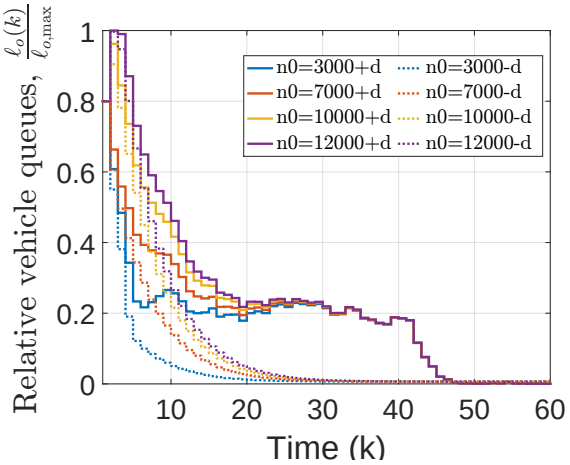


(f) Origin link 6

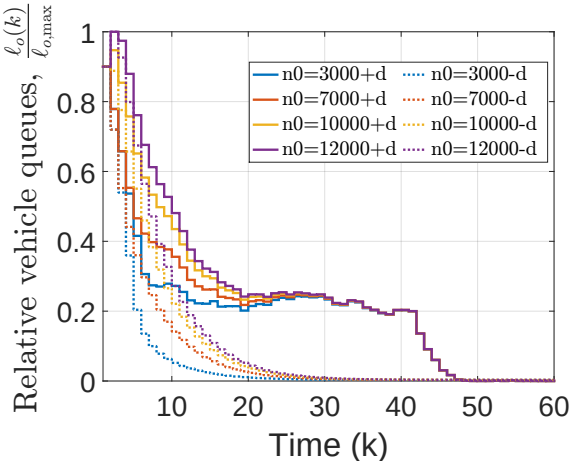
Figure 4.23: Medium demand scenario: Relative vehicle queues at origin links 1–6.



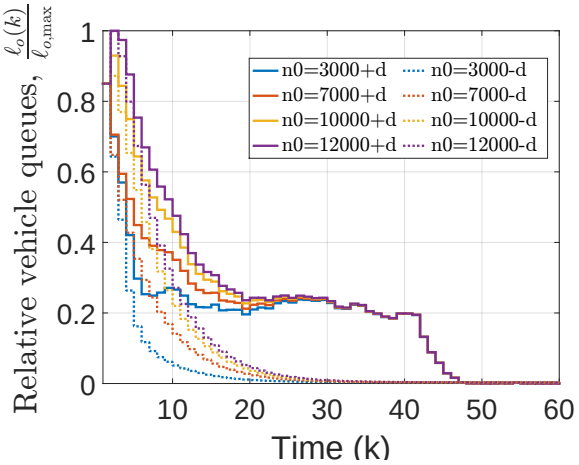
(a) Origin link 7



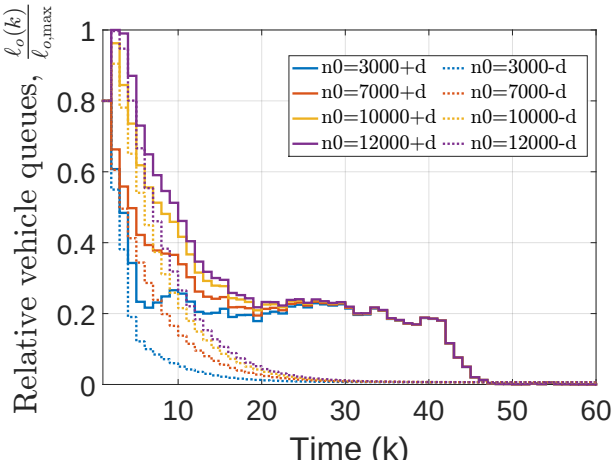
(b) Origin link 8



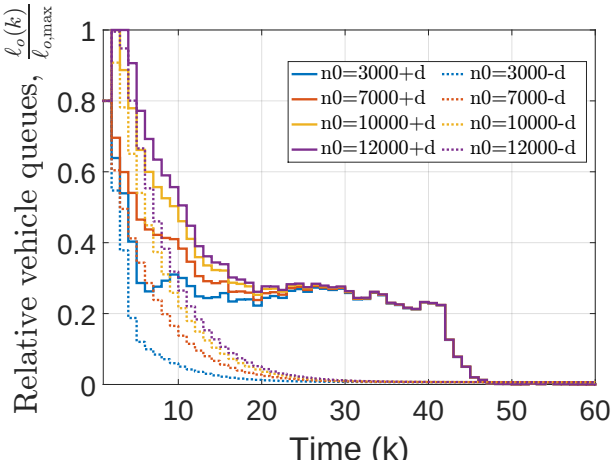
(c) Origin link 9



(d) Origin link 10

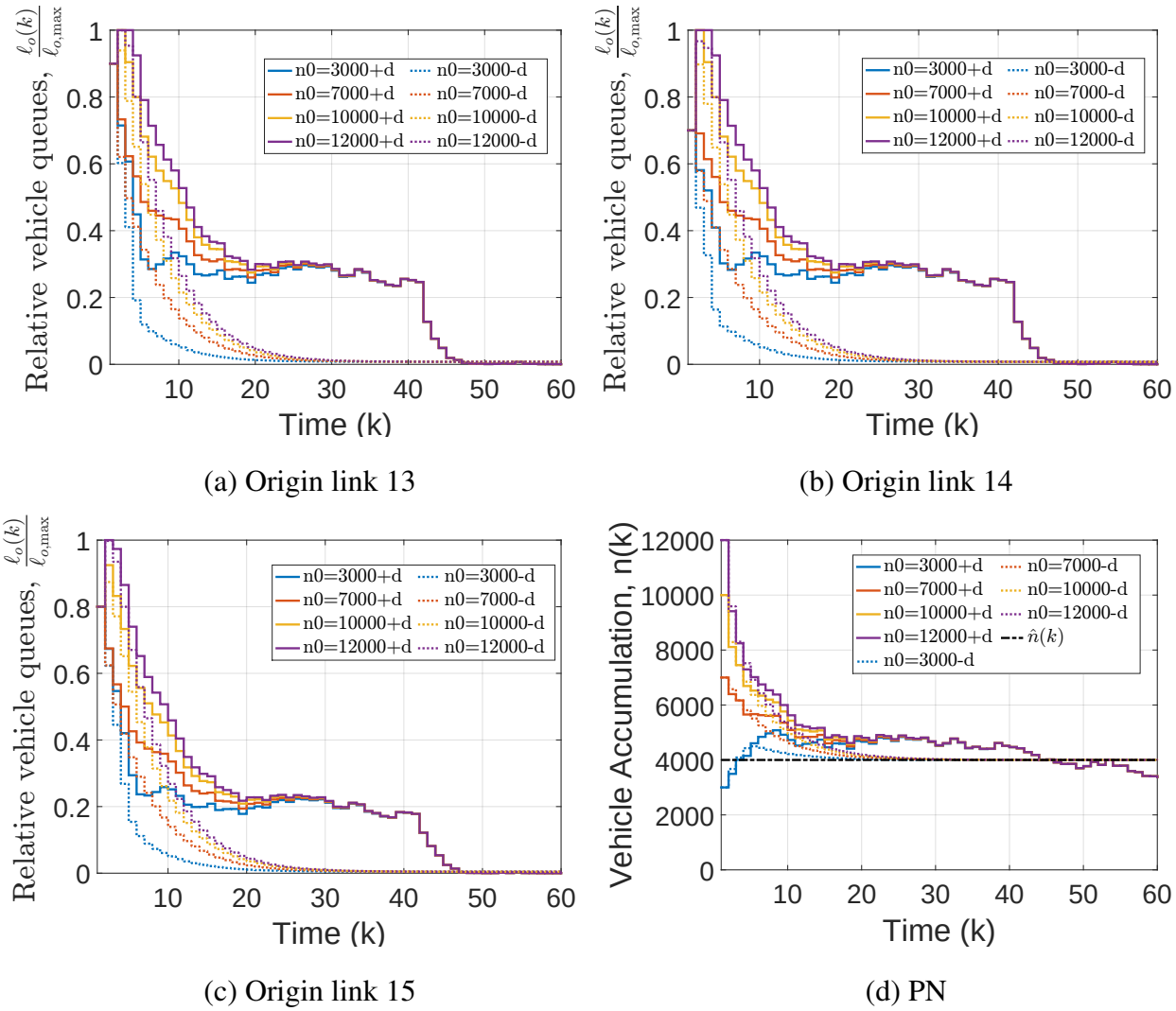


(e) Origin link 11



(f) Origin link 12

Figure 4.24: Medium demand scenario: Relative vehicle queues at origin links 7–12.



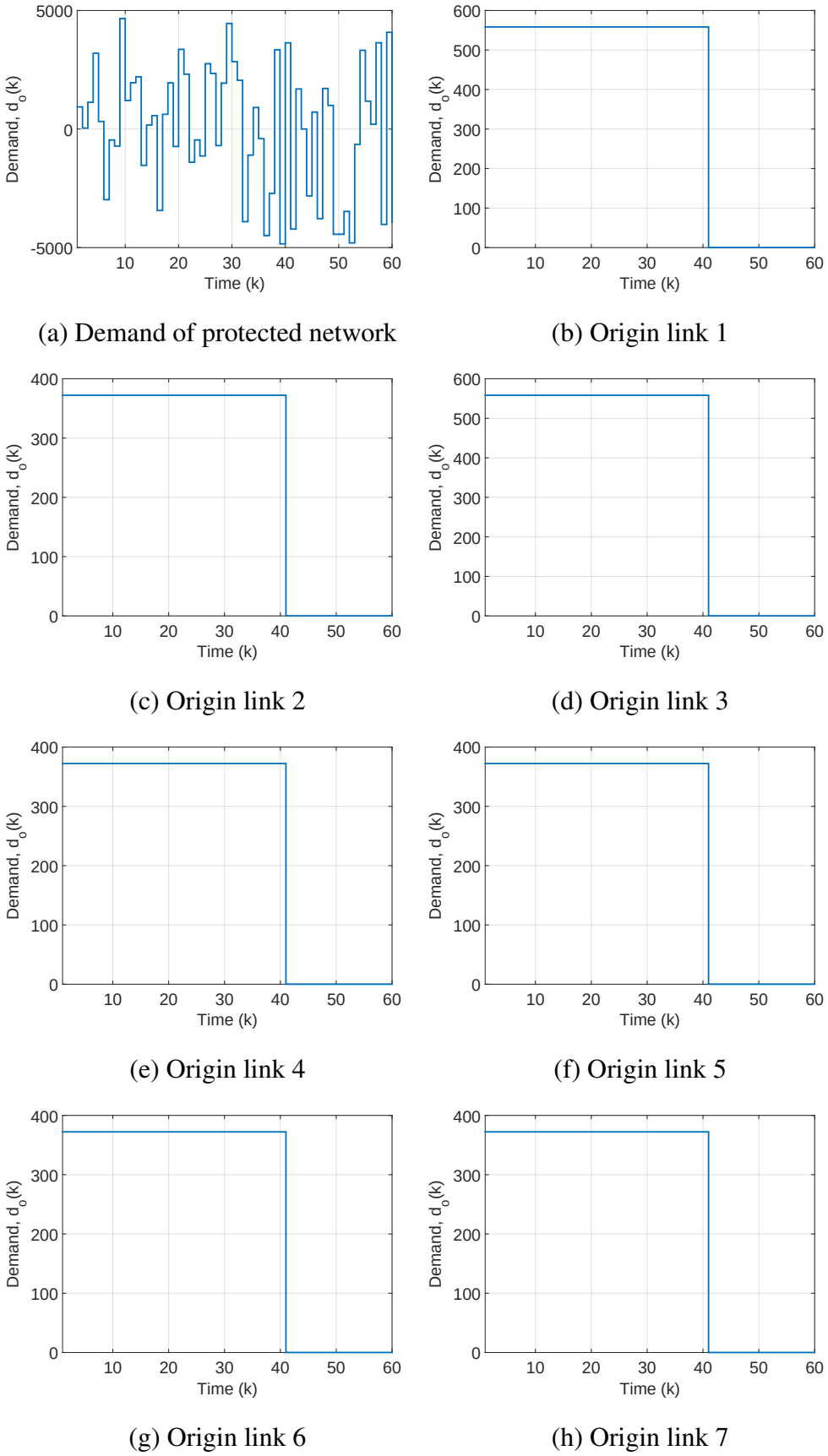
(a) Origin link 13

(b) Origin link 14

(c) Origin link 15

(d) PN

Figure 4.25: Medium demand scenario: (a)–(c) Relative vehicle queues at origin links 13–15; (d) State trajectory at PN.



(a) Demand of protected network

(b) Origin link 1

(c) Origin link 2

(d) Origin link 3

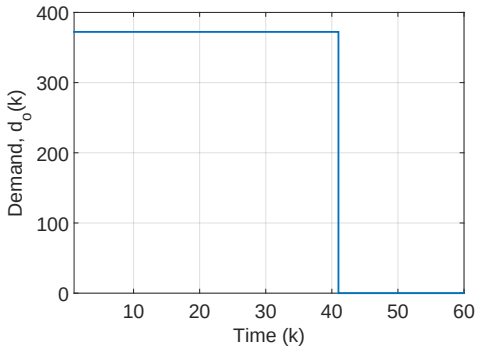
(e) Origin link 4

(f) Origin link 5

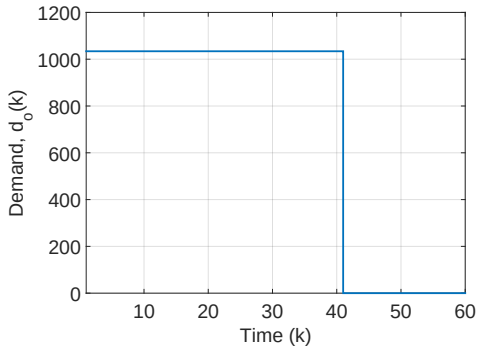
(g) Origin link 6

(h) Origin link 7

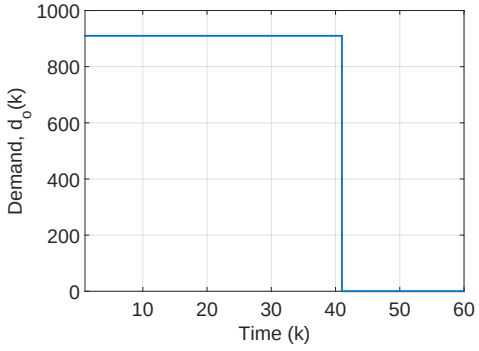
Figure 4.26: High demand scenario; (a): Demand at PN; (b)–(h) demand at origin links 1–7.



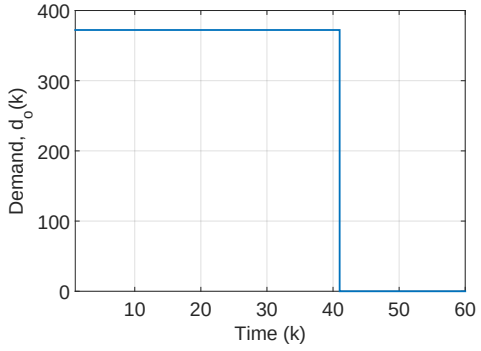
(a) Origin link 8



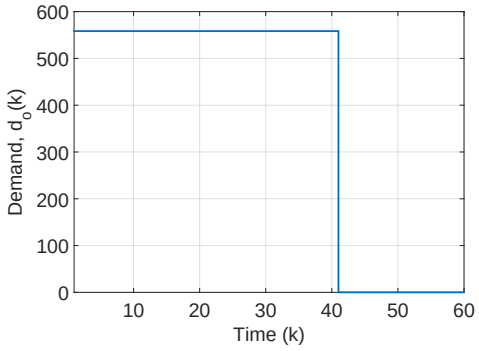
(b) Origin link 9



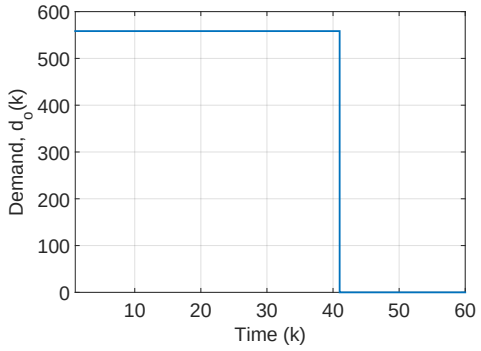
(c Origin link 10



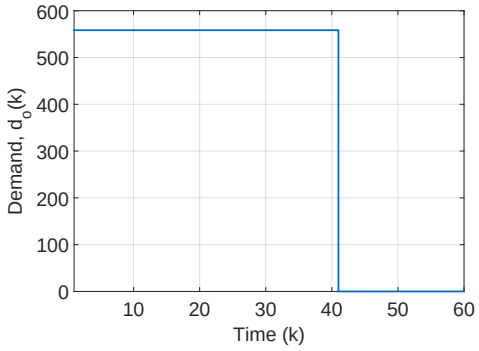
(d) Origin link 11



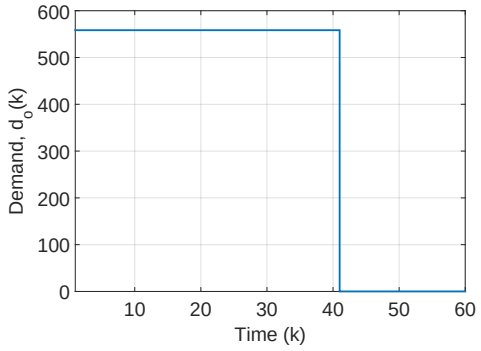
(e) Origin link 12



(f) Origin link 13

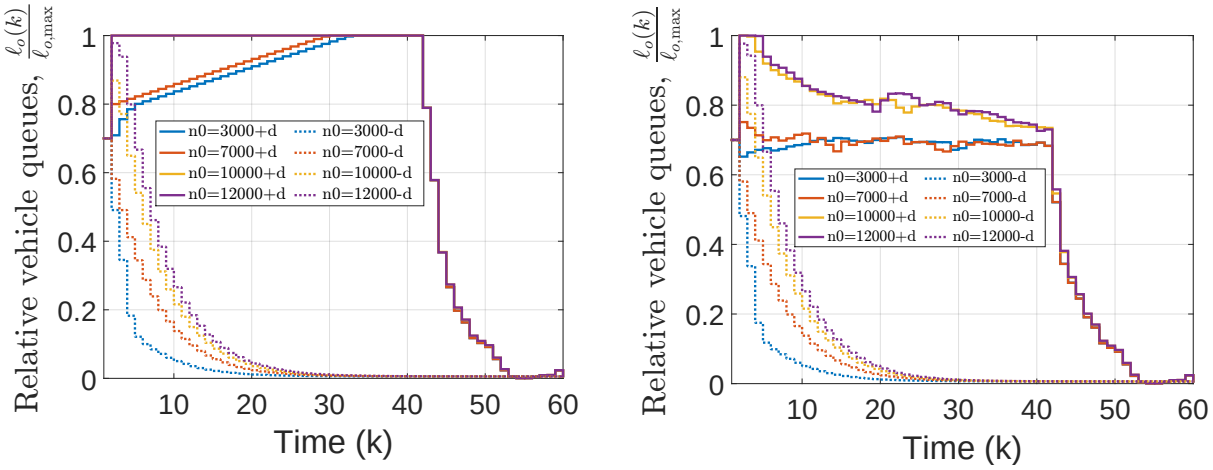


(g) Origin link 14



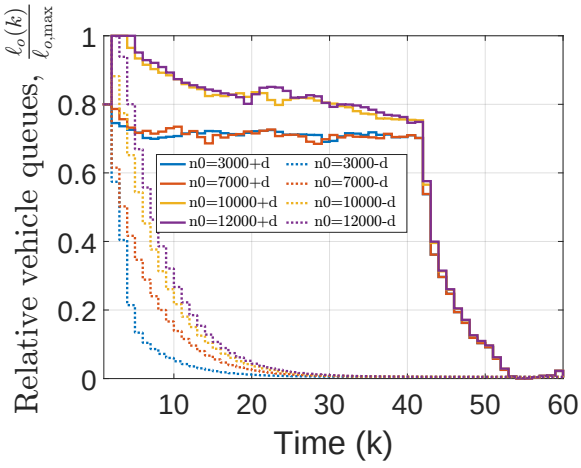
(h) Origin link 15

Figure 4.27: High demand scenario: Demand at origin links 8–15.

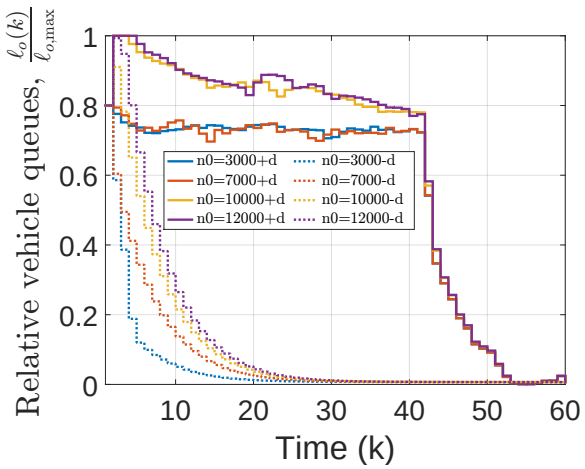


(a) Origin link 1

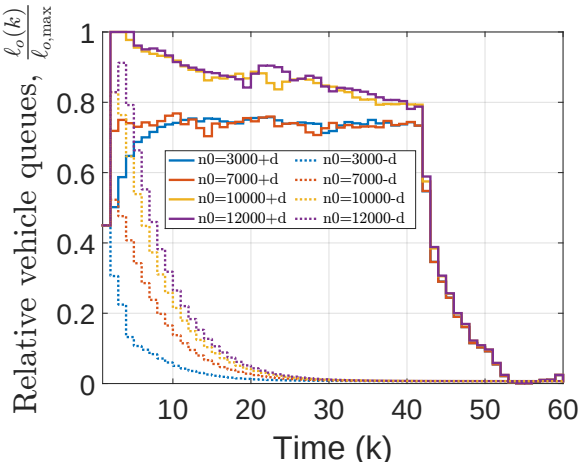
(b) Origin link 2



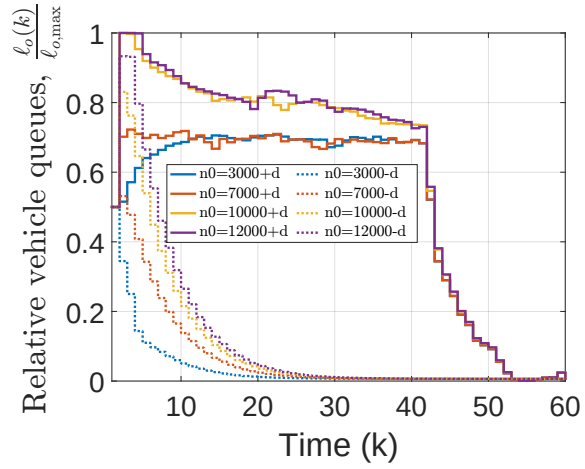
(c) Origin link 3



(d) Origin link 4

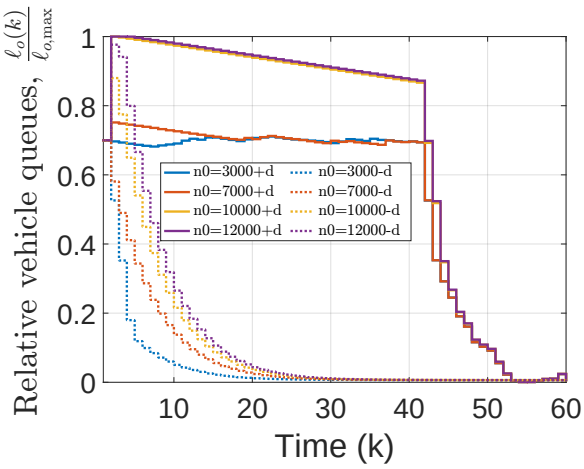


(e) Origin link 5

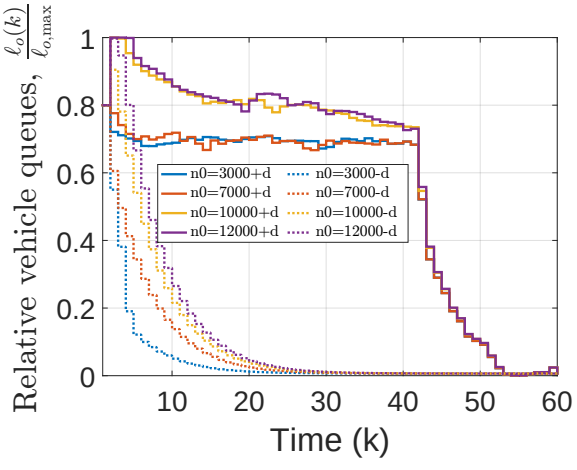


(f) Origin link 6

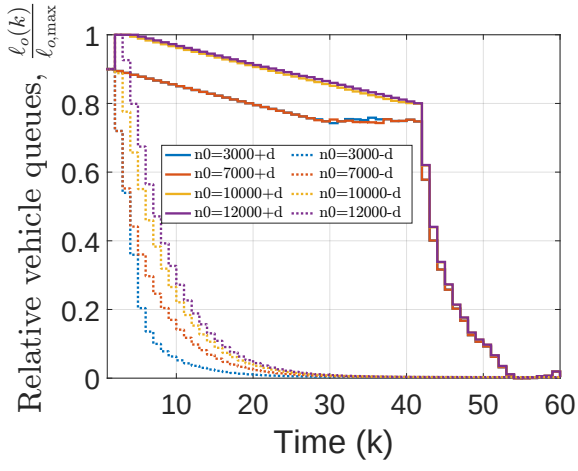
Figure 4.28: High demand scenario: Relative vehicle queues at origin links 1–6.



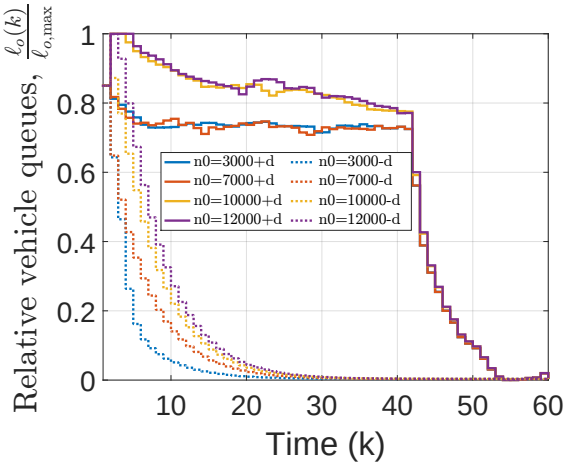
(a) Origin link 7



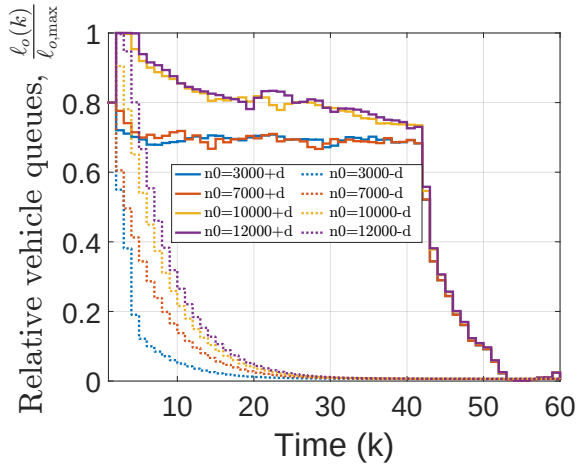
(b) Origin link 8



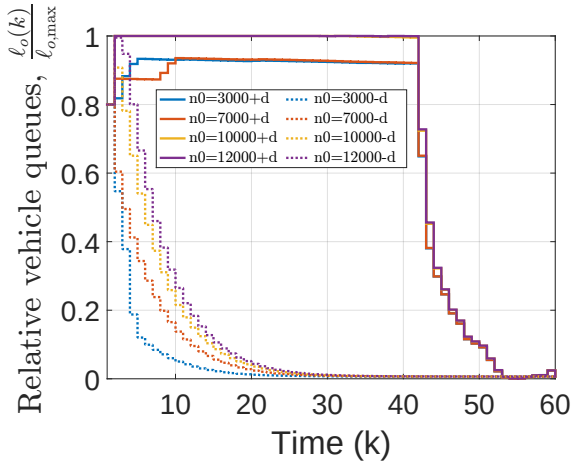
(c) Origin link 9



(d) Origin link 10

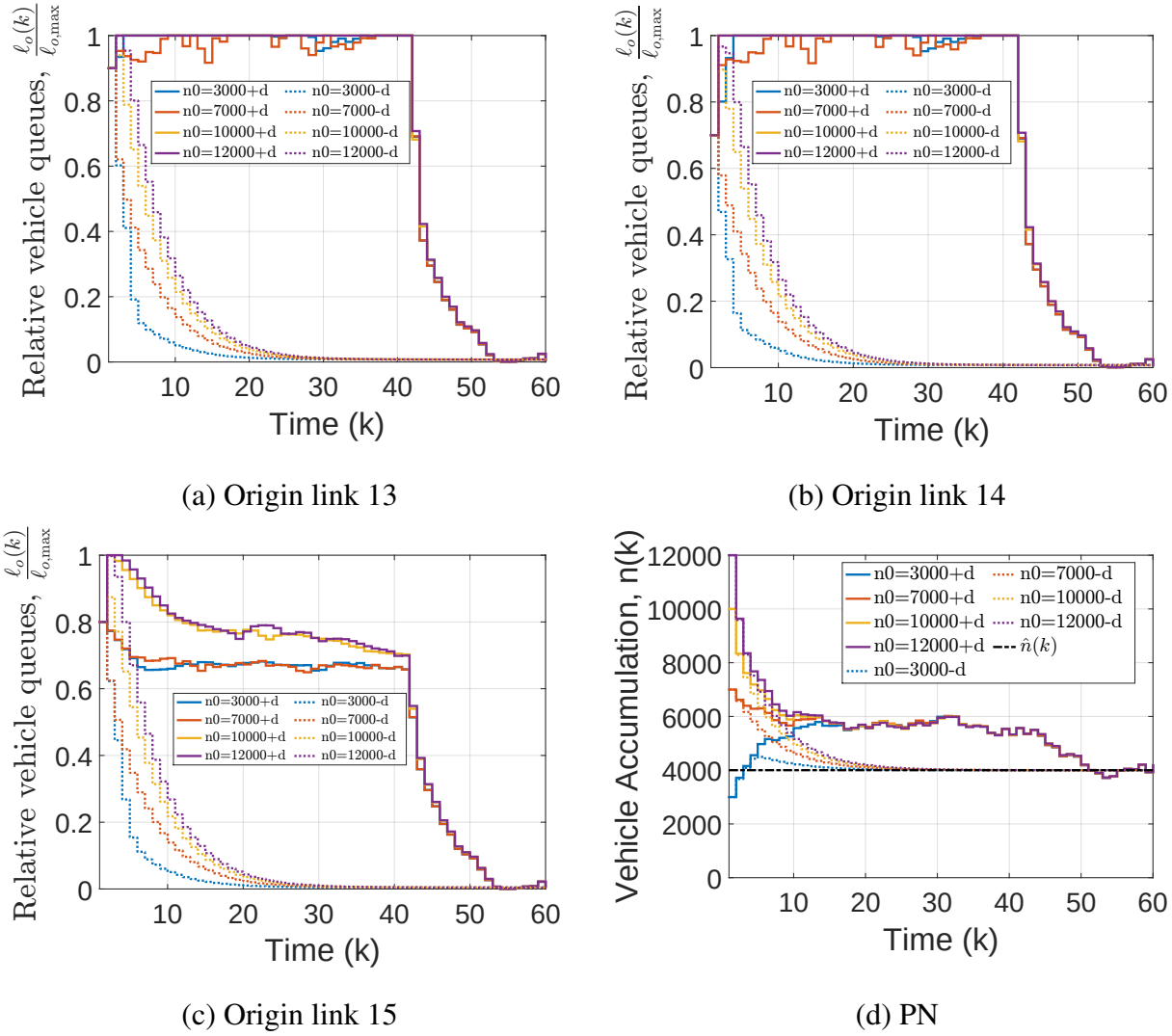


(e) Origin link 11



(f) Origin link 12

Figure 4.29: High demand scenario: Relative vehicle queues at origin links 7–12.



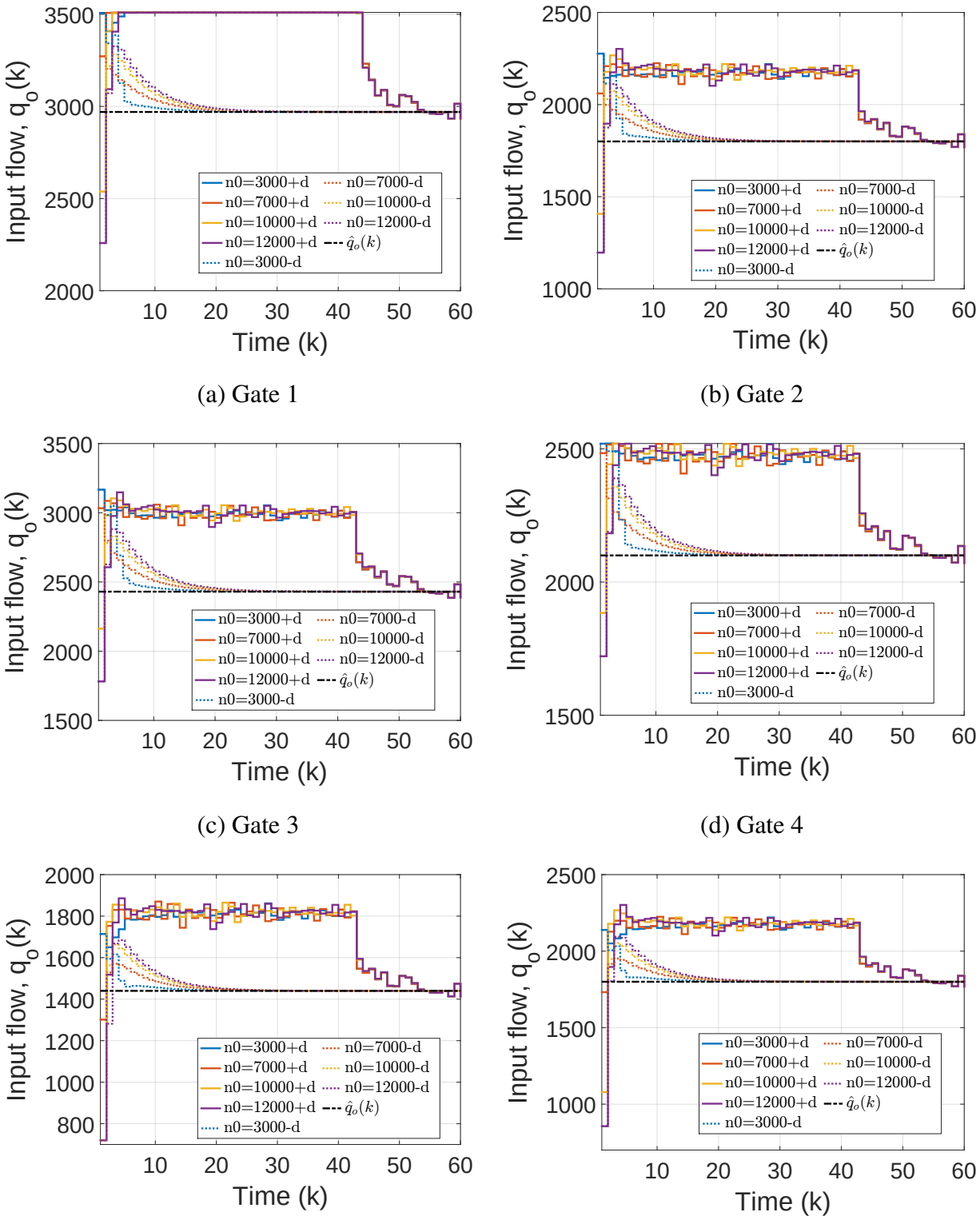
(a) Origin link 13

(b) Origin link 14

(c) Origin link 15

(d) PN

Figure 4.30: High demand scenario; (a)–(c): Relative vehicle queues at origin links 13–15; (d) State trajectory of PN.



(a) Gate 1

(b) Gate 2

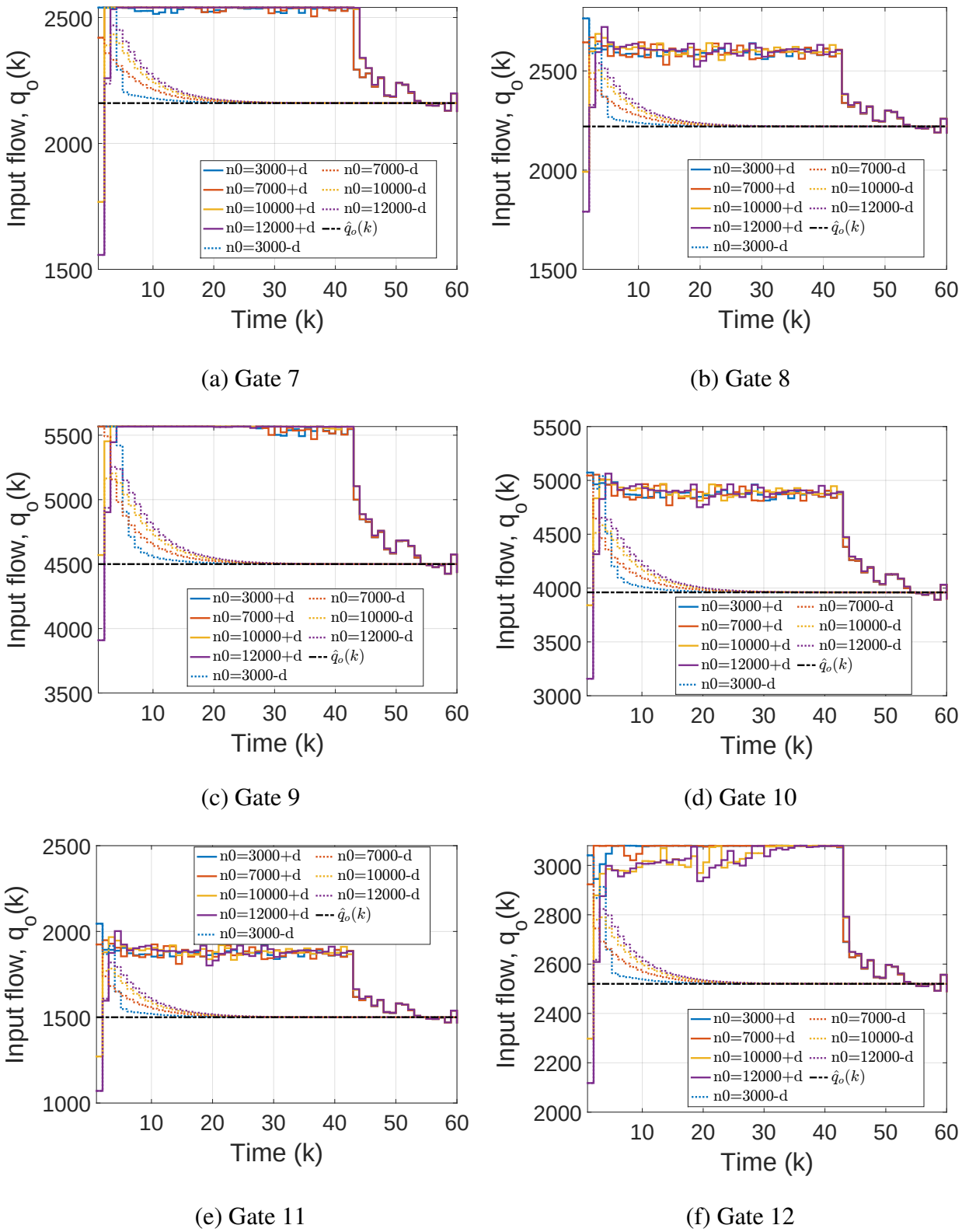
(c) Gate 3

(d) Gate 4

(e) Gate 5

(f) Gate 6

Figure 4.31: High demand scenario: Ordered flow at gates 1–6.



(a) Gate 7

(b) Gate 8

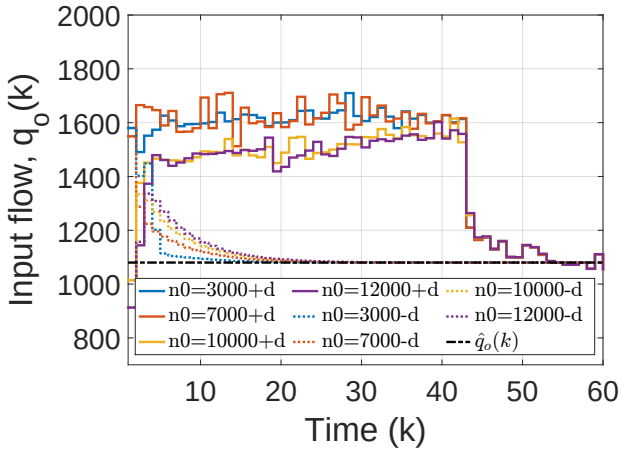
(c) Gate 9

(d) Gate 10

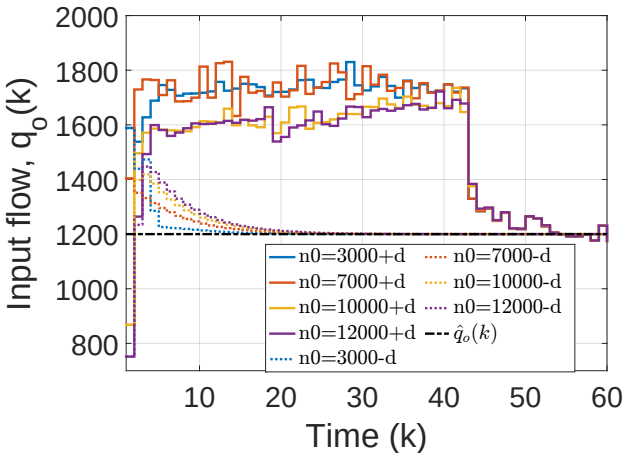
(e) Gate 11

(f) Gate 12

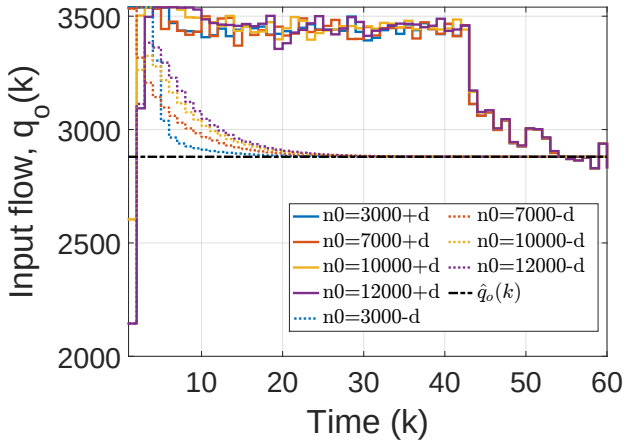
Figure 4.32: High demand scenario: Ordered flow at gates 7–12.



(a) Gate 13



(b) Gate 14



(c) Gate 15

Figure 4.33: High demand scenario: Ordered flow at gates 13–15.

4.5.3.4 Perimeter-ordered flow allocation policies: CAP & OAP

This section investigates the performance of the two perimeter-ordered flow allocation policies presented in Section 4.4, namely capacity-based flow allocation policy (CAP) and optimisation-based flow allocation policy (OAP). These policies are then compared with the proposed multi-gated perimeter control in the next section. The main task of these policies is to allocate a global perimeter-ordered flow to a number of candidate gates/junctions at the periphery of the network by taking into account the different geometric characteristics of origin links, i.e., length, number of lanes, storage capacity, etc. Several tests were conducted under different scenarios in order to investigate the behaviour of the proposed allocation policies. For comparable results, the characteristics of each scenario and demand were similar to the medium demand in Section 4.5.3.3.

Capacity-based flow allocation policy (CAP)

Table 4.5 provides the storage capacity-based ratios of the fifteen entrance links. These ratios were used to allocate a global perimeter-ordered flow (here Version II is used with $R = 0.00001$) to the fifteen gates/junctions at the periphery of the network. It is evident that links with the same storage capacity possess the same ratios.

Figures 4.39–4.41 and 4.42–4.44 show the control trajectories (global input flow distribution) and relative queues at origin links. CAP is seen to distribute the global flow close to its reference point $\hat{\mathbf{q}}$ of gating links at $k \approx 20$. However, there are some gates with significant differences between the ordered input flow and desired point (see e.g., Figure 4.39(a) and 4.39(d)), indicating a drawback of the strategy.

It can be seen that CAP needs more time to dissolve the relative vehicle queues developed at the periphery of the network (Figures 4.42– 4.44). However, under CAP, a high number of trips is seen to be unserved at the end of simulation. Having some trips unserved allowed CAP to secure the protected network area from overloading (Figure 4.44(d)). Stabilisation around the reference state $\hat{n} = 4000$ veh is seen at $k = 20$. The short settling time is attributed to the complete lack of information of the developed queues at origin links.

Optimisation-based flow allocation policy (OAP)

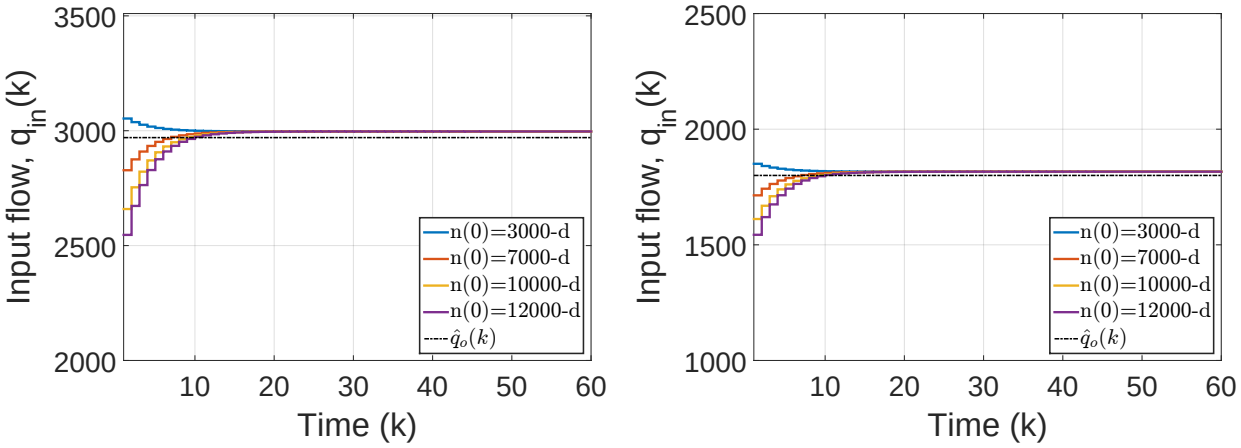
Figures 4.34–4.35 and 4.36–4.38 show the control trajectories (global input flow distribution) and relative queues at origin links. As can be seen OAP provides similar performance to CAP. Again there are some gates with significant differences between the ordered input flow and desired point (see e.g., Figures 4.34(a), 4.34(d), and 4.34(f)).

OAP is also seen slow to dissolve the developed queues at the origin links. Moreover, relative vehicle queues remained unbalanced. For instance, link 11 and 12 simulated with the same $\ell(0)$, but link 12 experienced saturation for quite long time ($k > 40$). Finally stabilisation around

Table 4.5: Storage capacity-based ratios of the fifteen entrance links.

Gate	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Capacity	128	109	163	98	95	109	109	109	296	269	109	103	84	84	186
Ratio (%)	6.2	5.3	7.9	4.8	4.6	5.3	5.3	5.3	14.4	13.1	5.3	5	4.1	4.1	9.1

the reference state $\hat{n} = 4000$ veh is seen at $k = 20$ as in CAP. Concluding CAP and OAP indicated more or less the same performance for the considered simulation scenarios. High fidelity simulations in a microscopic environment would provide additional evidence on their performance under different networks and scenarios.



(a) Gate 1

(b) Gate 2

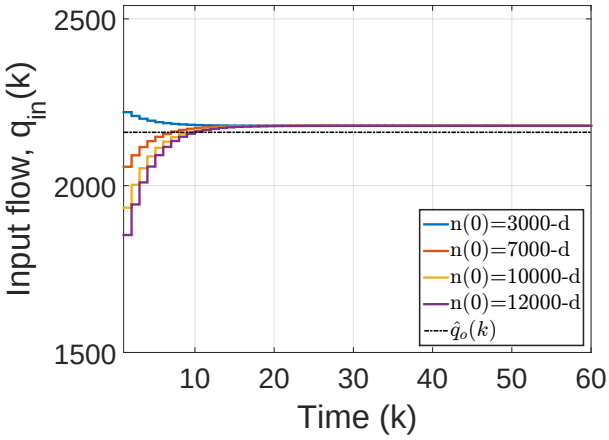
(c) Gate 3

(d) Gate 4

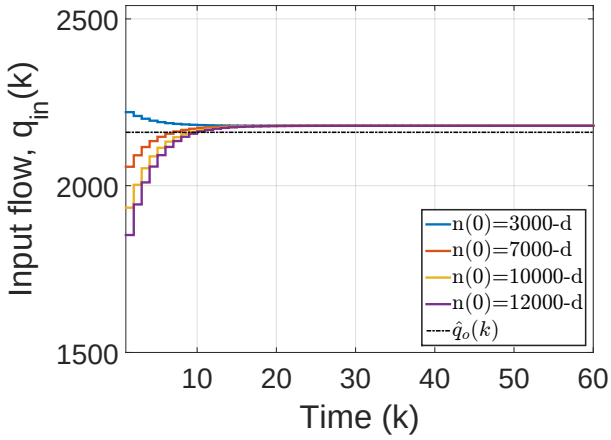
(e) Gate 5

(f) Gate 6

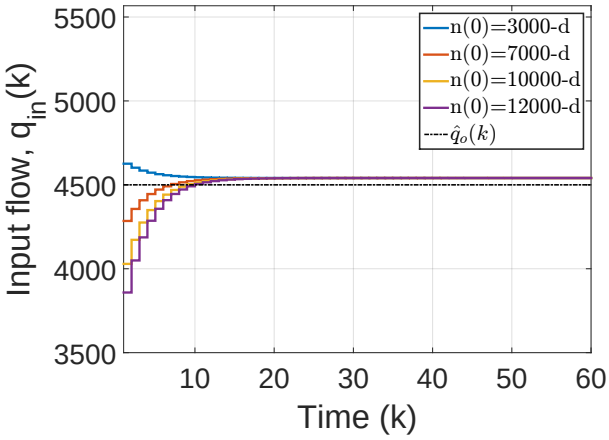
Figure 4.34: OAP: Ordered flow at gates 1–6.



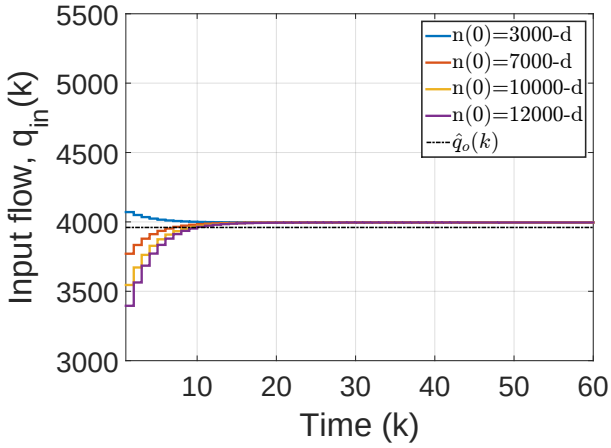
(a) Gate 7



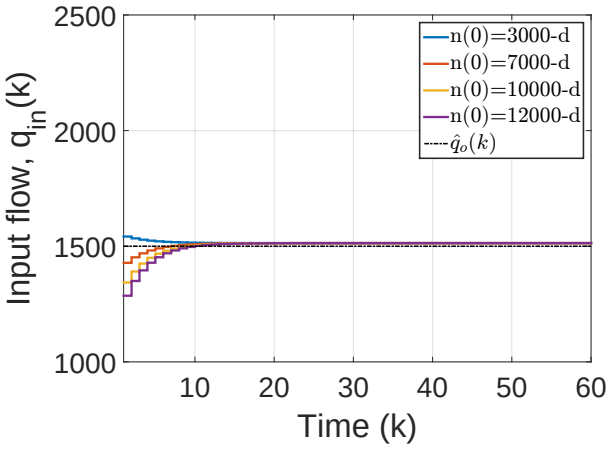
(b) Gate 8



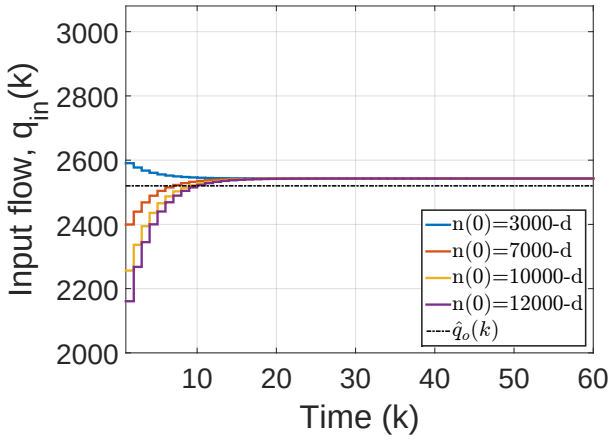
(c) Gate 9



(d) Gate 10

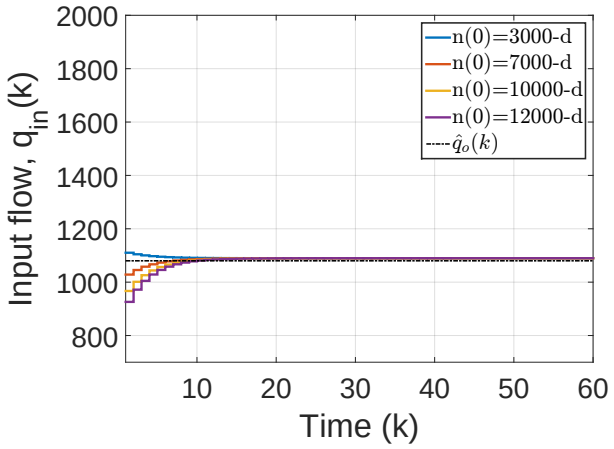


(e) Gate 11

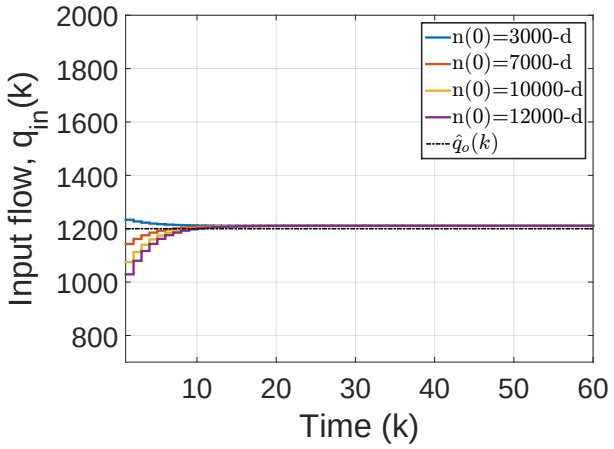


(f) Gate 12

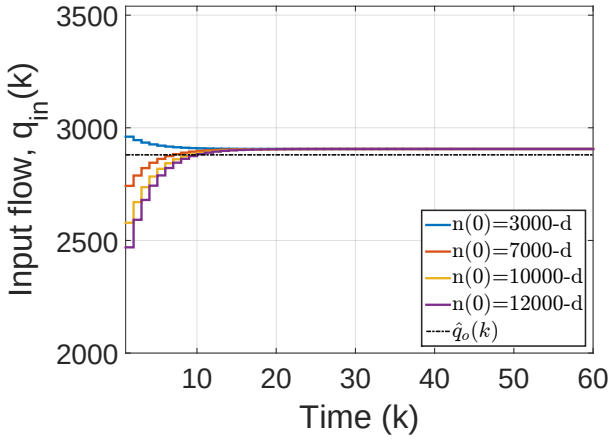
Figure 4.35: OAP: Ordered flow at gates 7–12.



(a) Gate 13

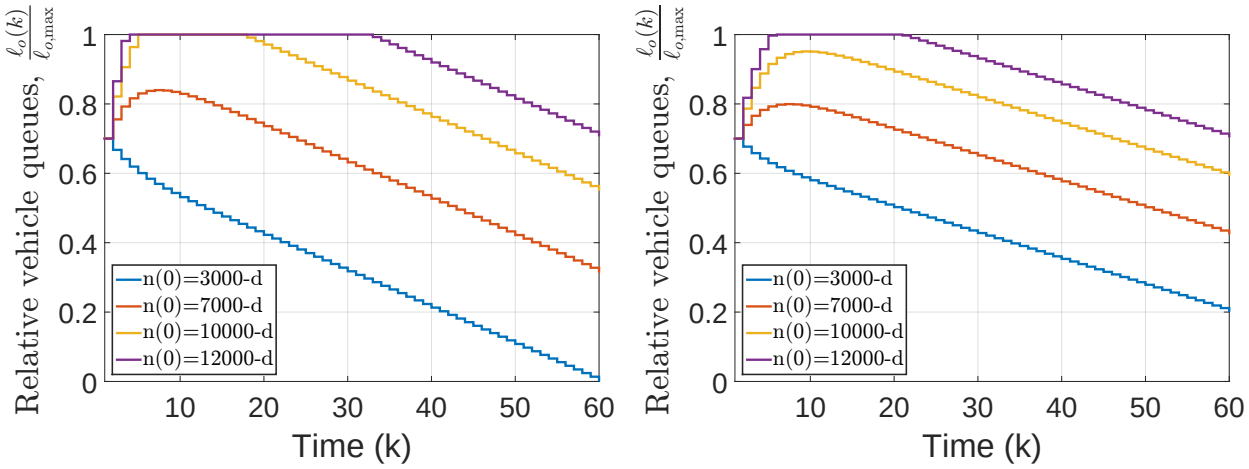


(b) Gate 14



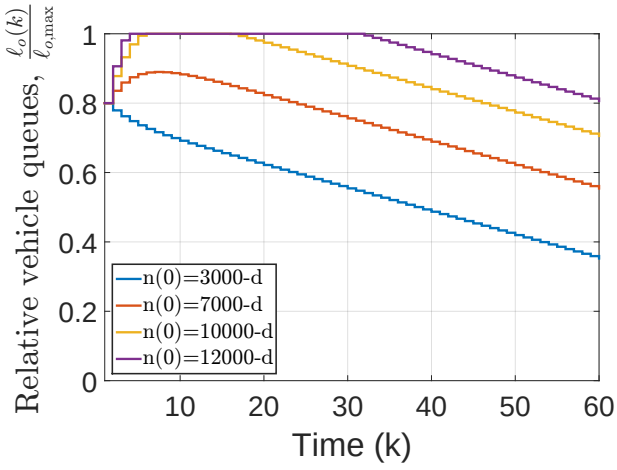
(c) Gate 15

Figure 4.35: OAP: Ordered flow at gates 13–15.

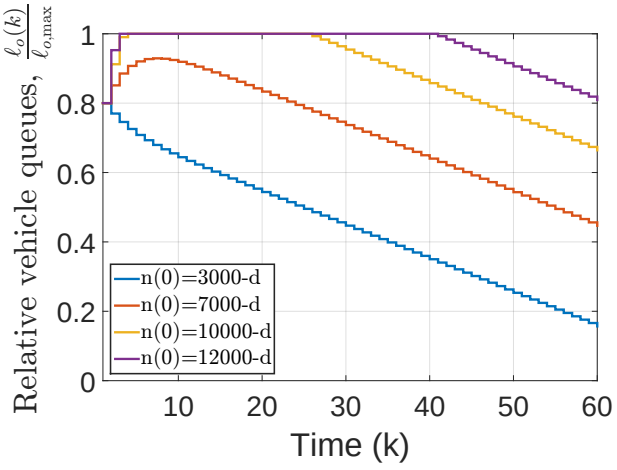


(a) Link 1

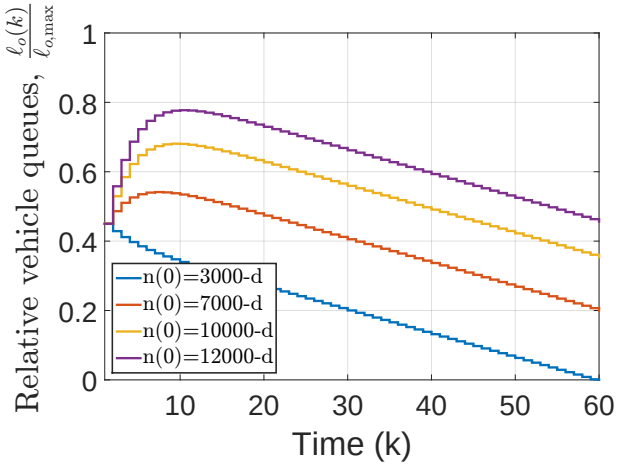
(b) Link 2



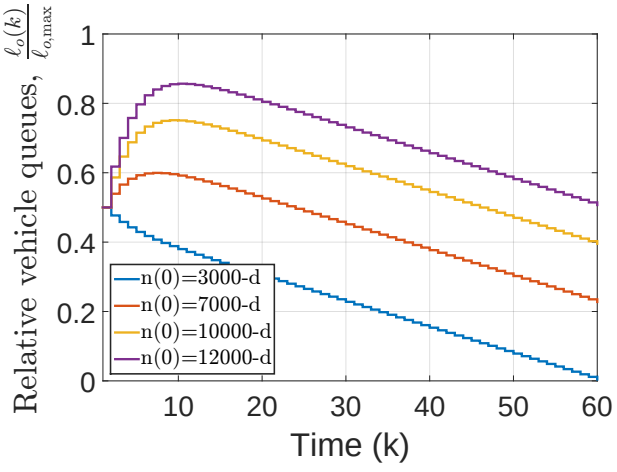
(c) Link 3



(d) Link 4

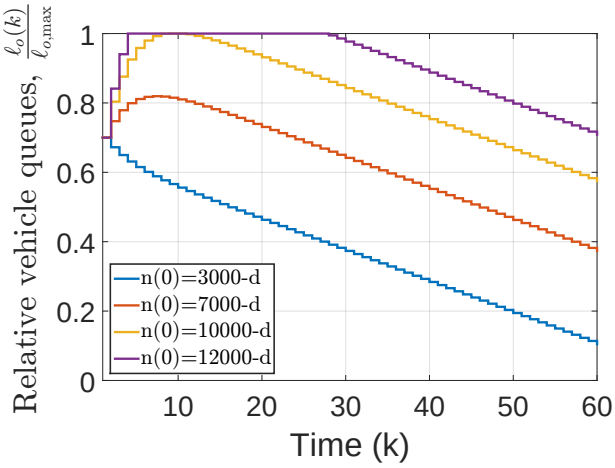


(e) Link 5

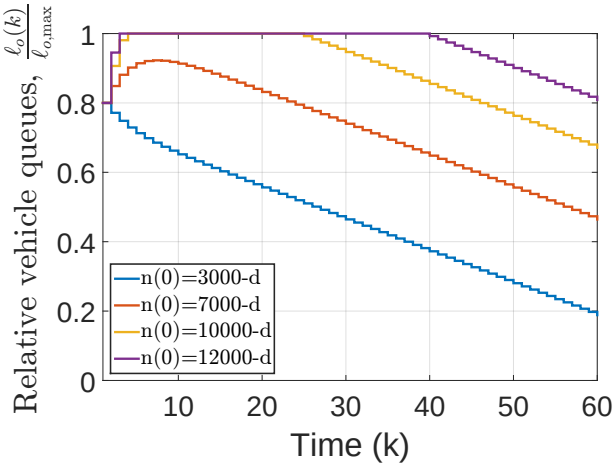


(f) Link 6

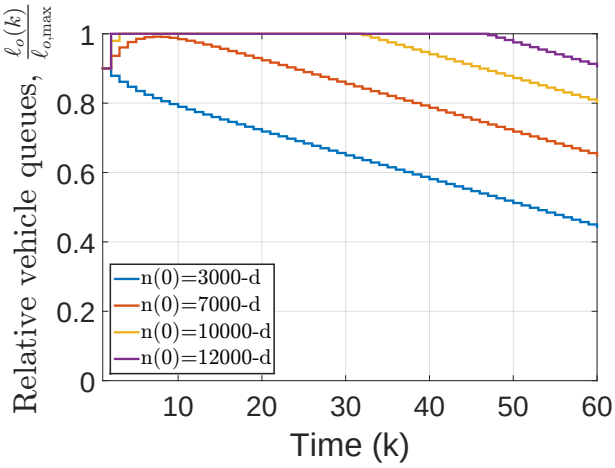
Figure 4.36: OAP: Relative vehicle queues at links 1–6.



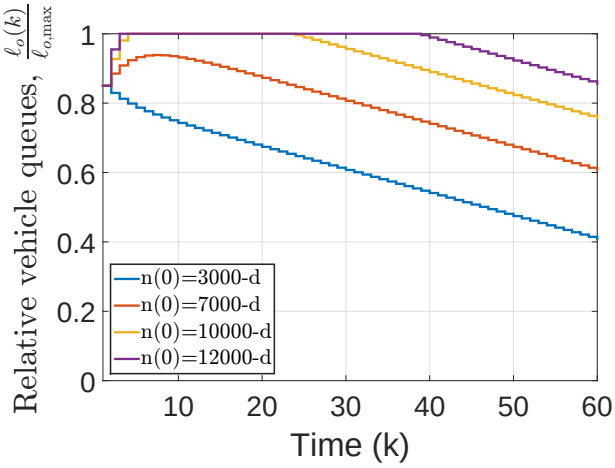
(a) Link 7



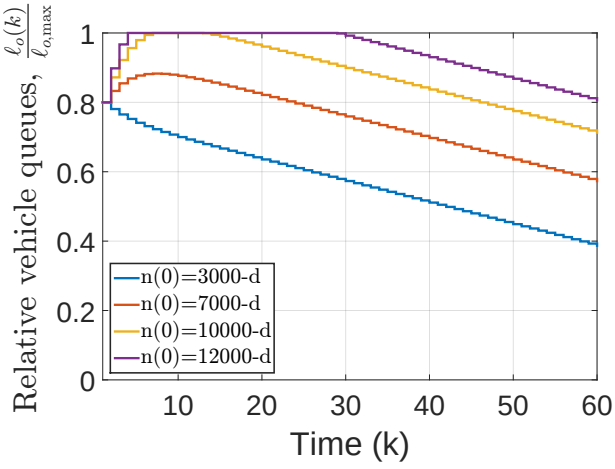
(b) Link 8



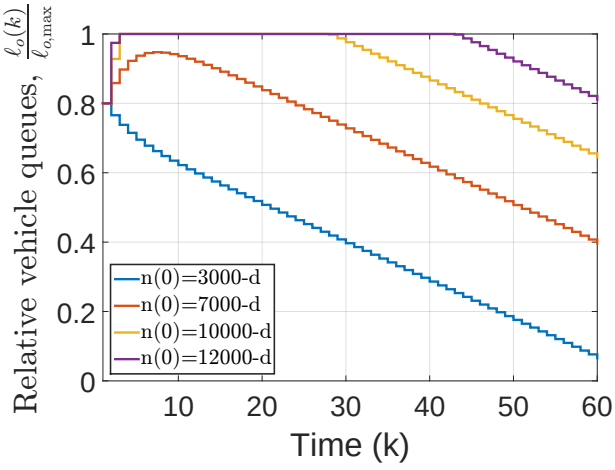
(c) Link 9



(d) Link 10

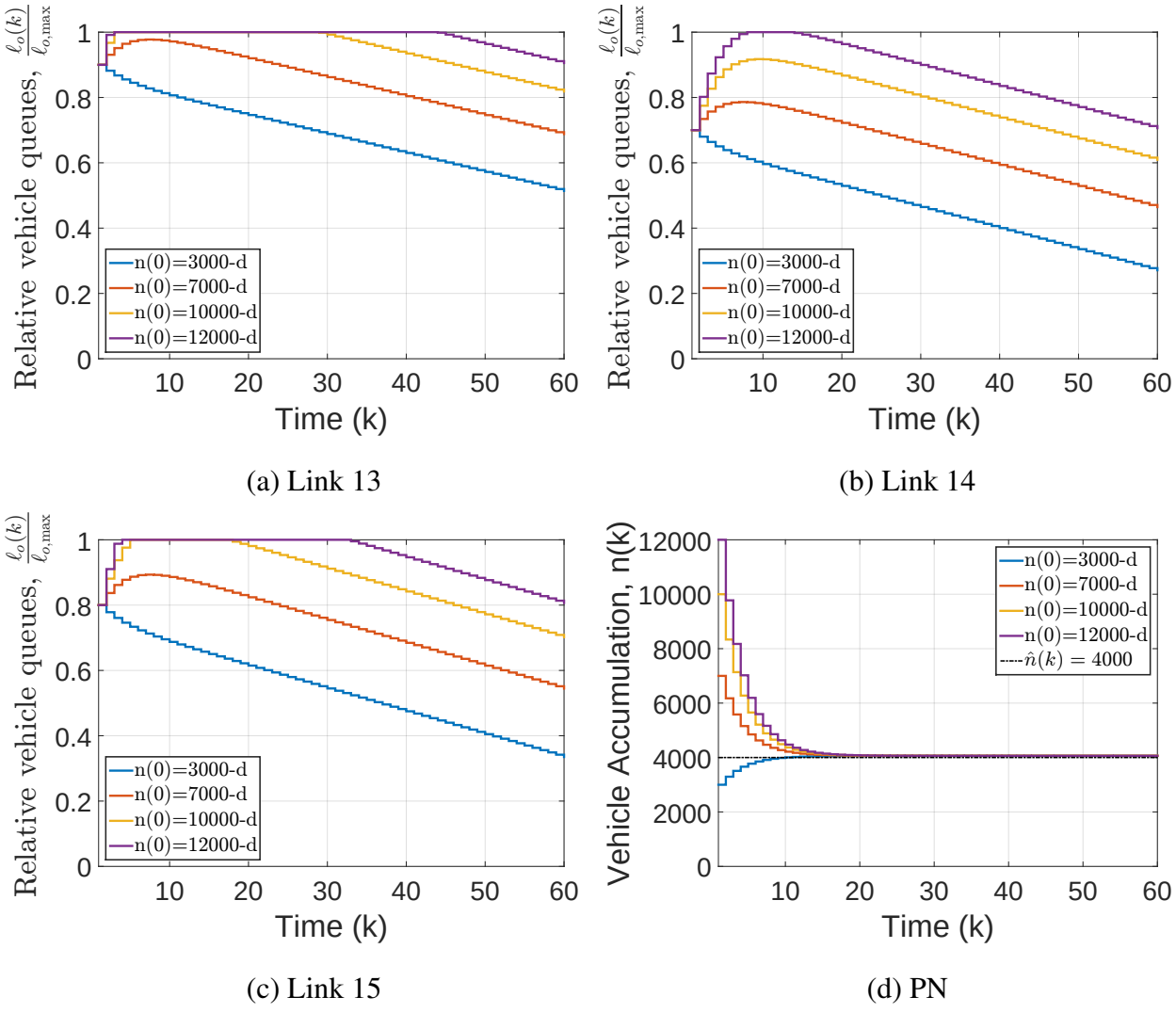


(e) Link 11



(f) Link 12

Figure 4.37: OAP: Relative vehicle queues at links 7–12.



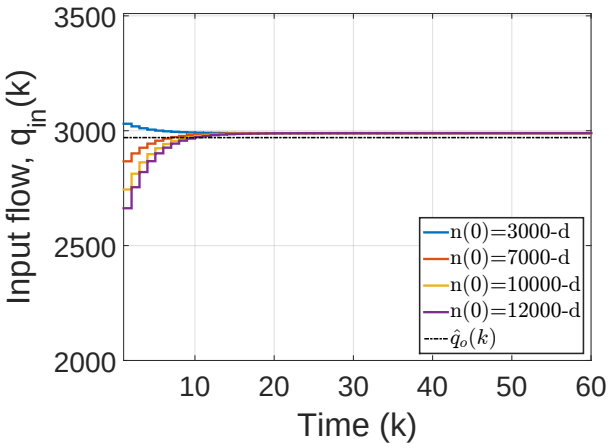
(a) Link 13

(b) Link 14

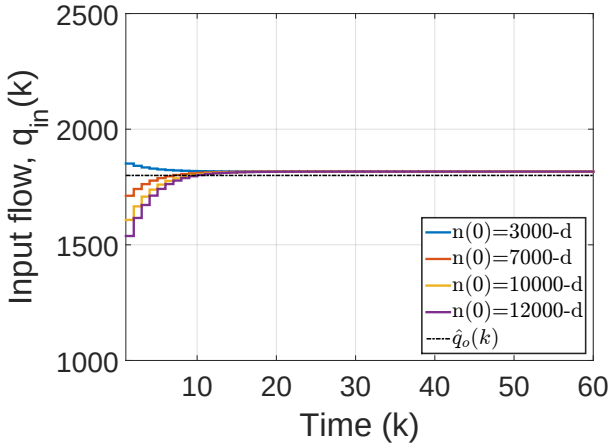
(c) Link 15

(d) PN

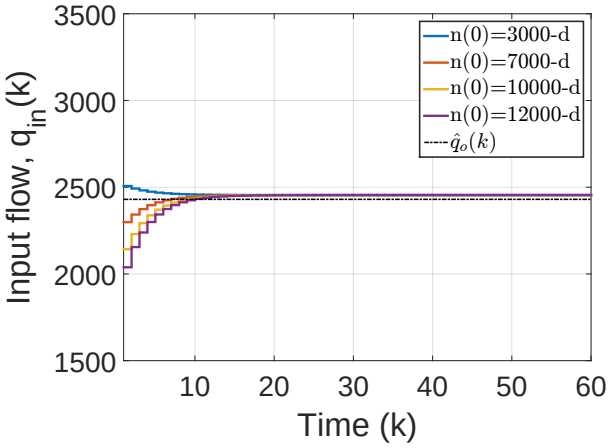
Figure 4.38: OAP: (a)–(c) Relative vehicle queues at links 13–15; (d) State trajectory at PN.



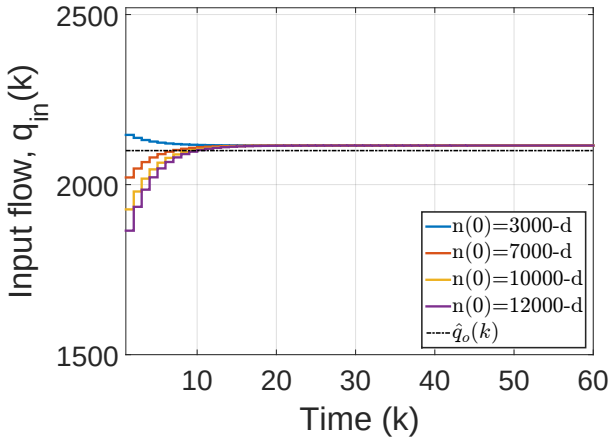
(a) Gate 1



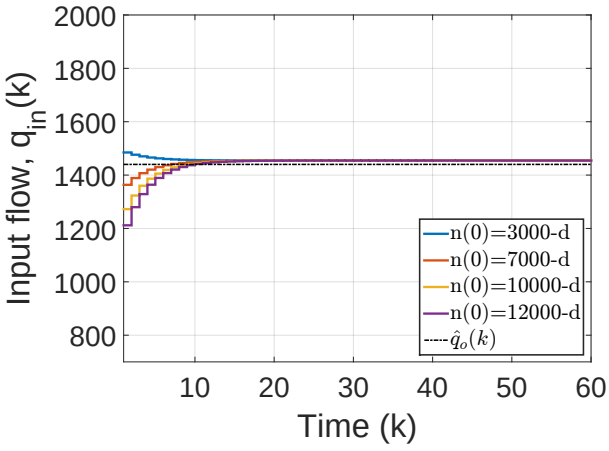
(b) Gate 2



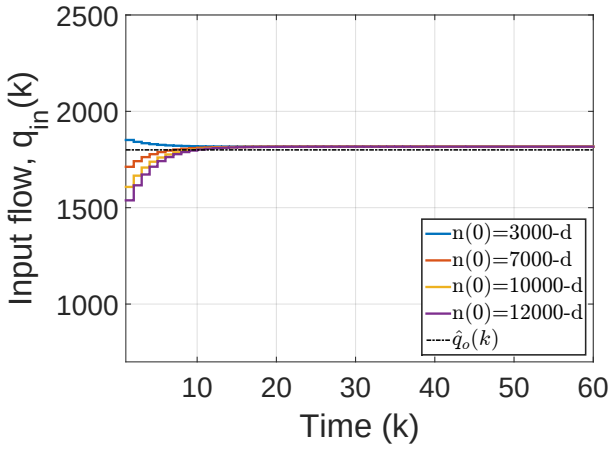
(c) Gate 3



(d) Gate 4

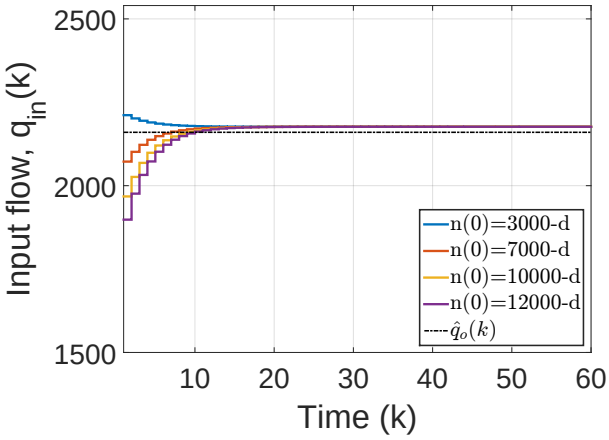


(e) Gate 5

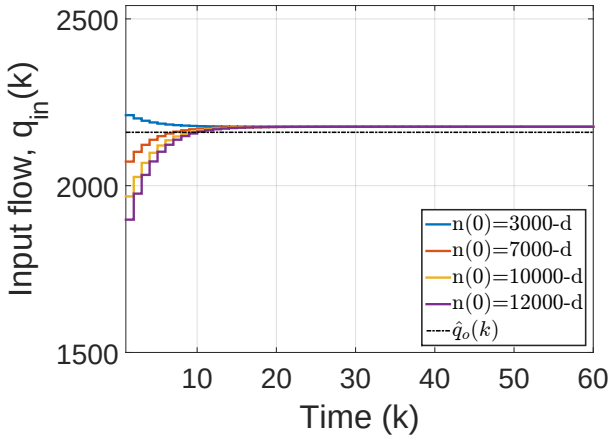


(f) Gate 6

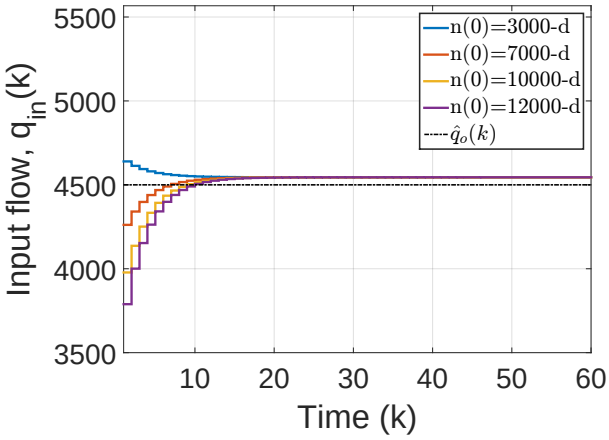
Figure 4.39: CAP: Ordered flow at gates 1–6.



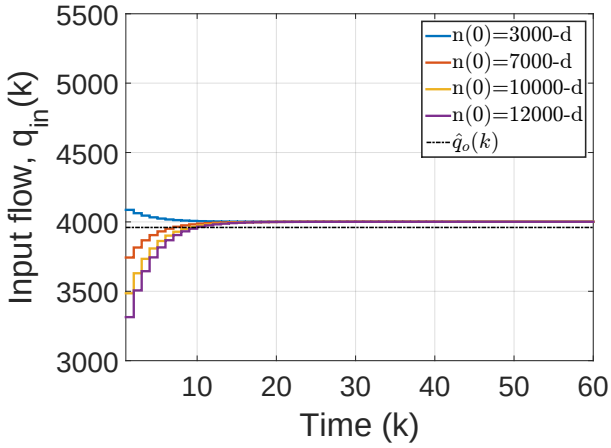
(a) Gate 7



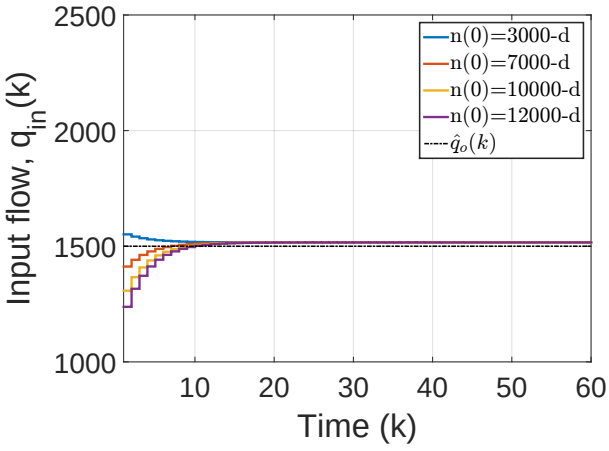
(b) Gate 8



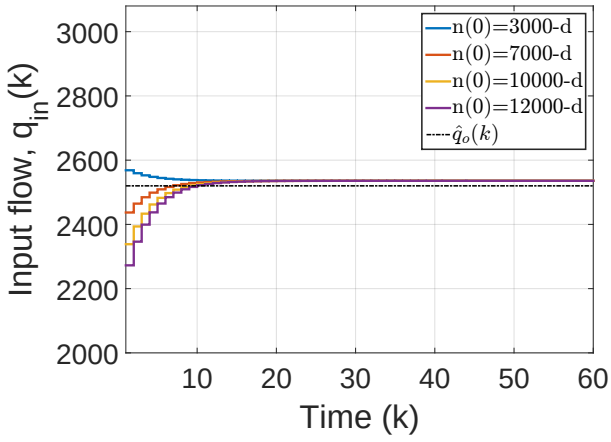
(c) Gate 9



(d) Gate 10

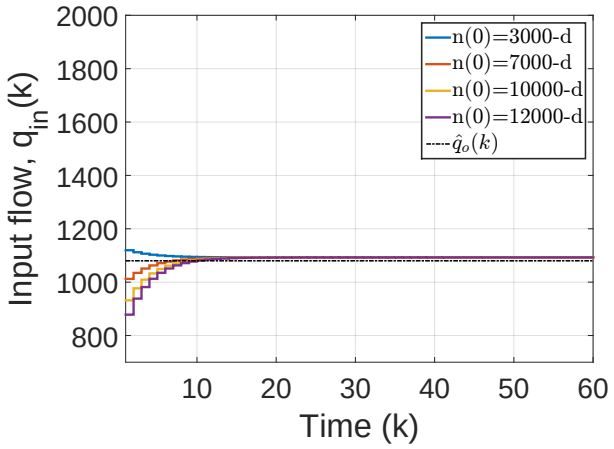


(e) Gate 11

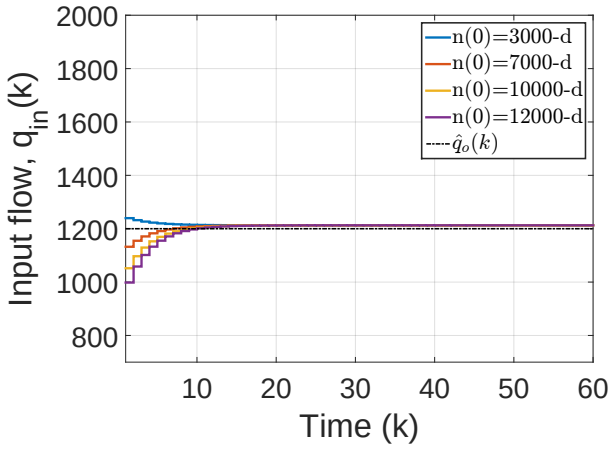


(f) Gate 12

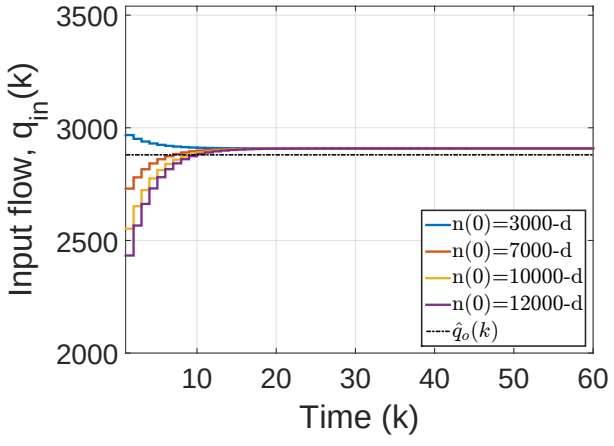
Figure 4.40: CAP: Ordered flow at gates 7–12.



(a) Gate 13

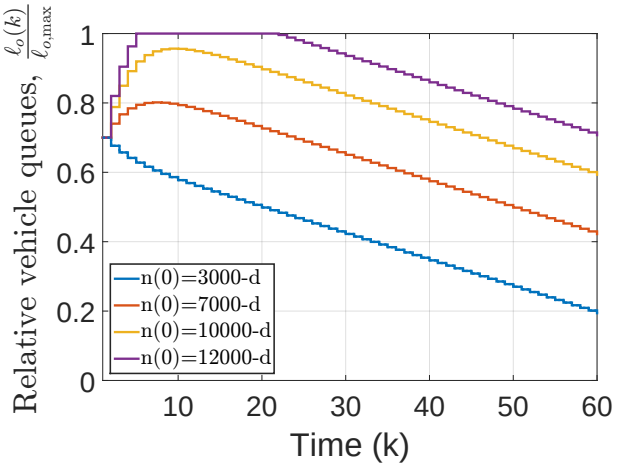


(b) Gate 14

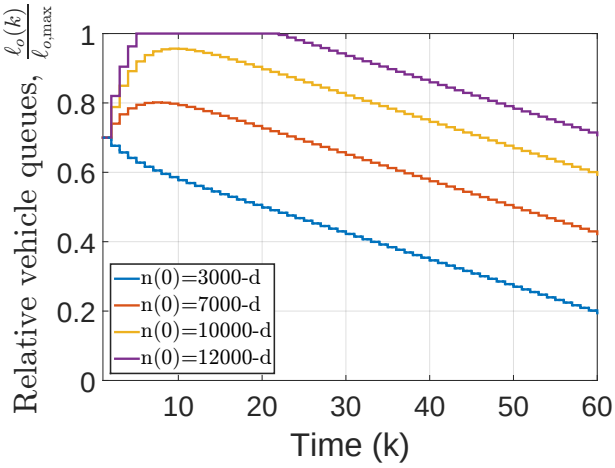


(c) Gate 15

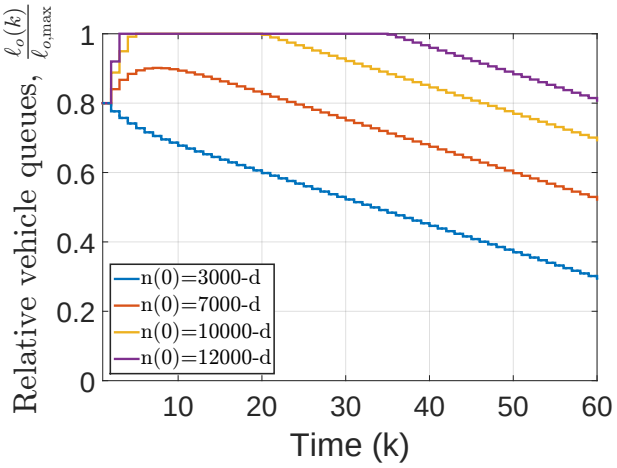
Figure 4.41: CAP: Ordered flow at gates 13–15.



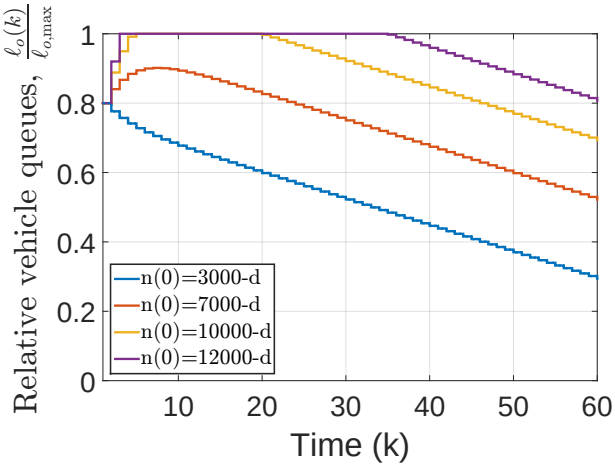
(a) Link 1



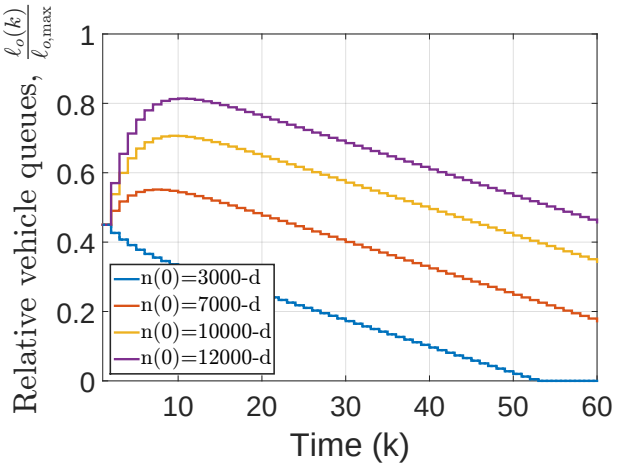
(b) Link 2



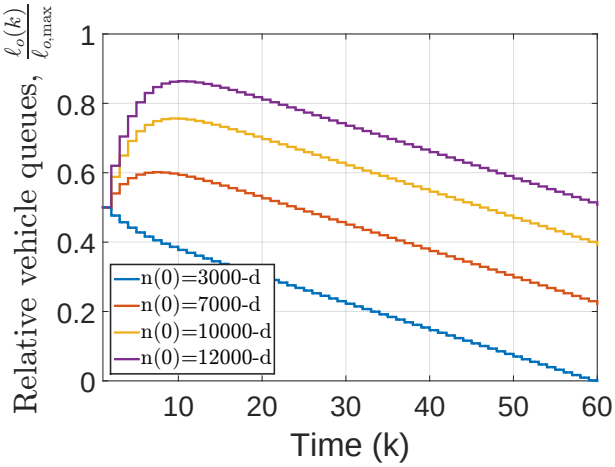
(c) Link 3



(d) Link 4

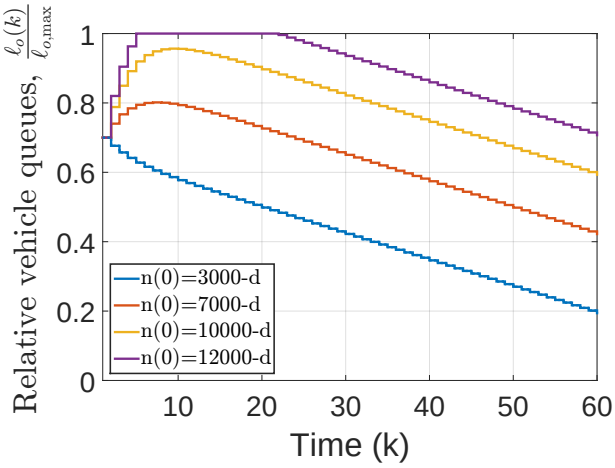


(e) Link 5

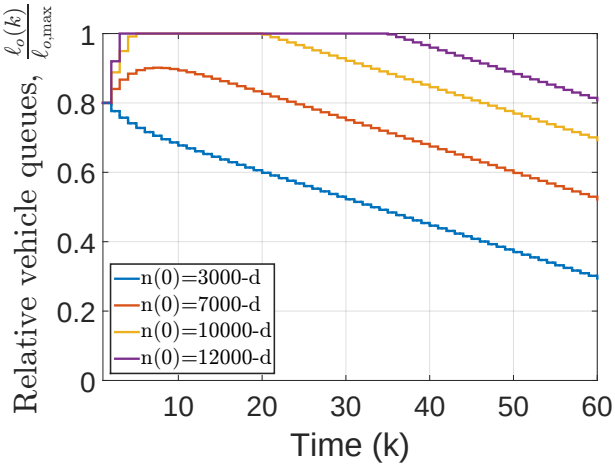


(f) Link 6

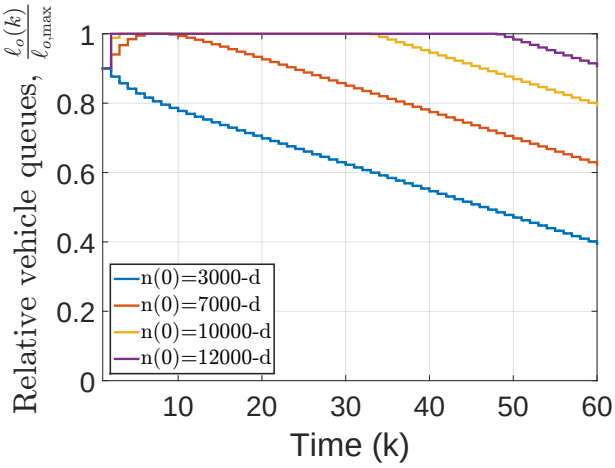
Figure 4.42: CAP: Relative vehicle queues at links 1–6.



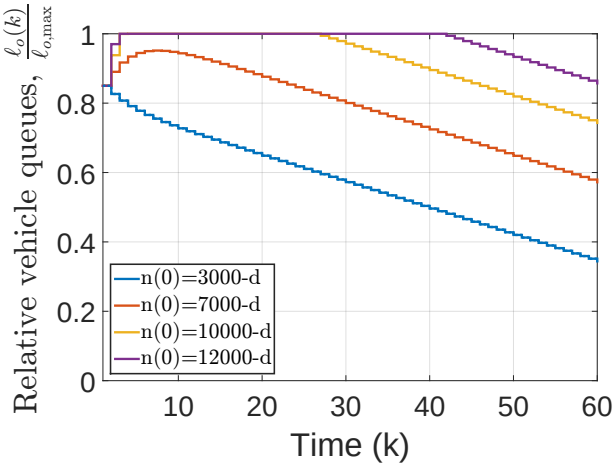
(a) Link 7



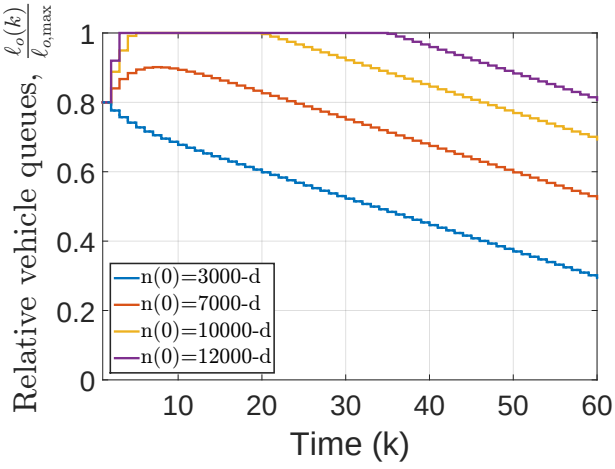
(b) Link 8



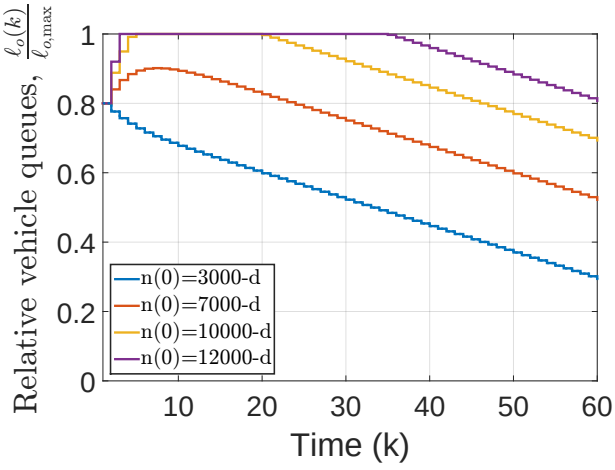
(c) Link 9



(d) Link 10

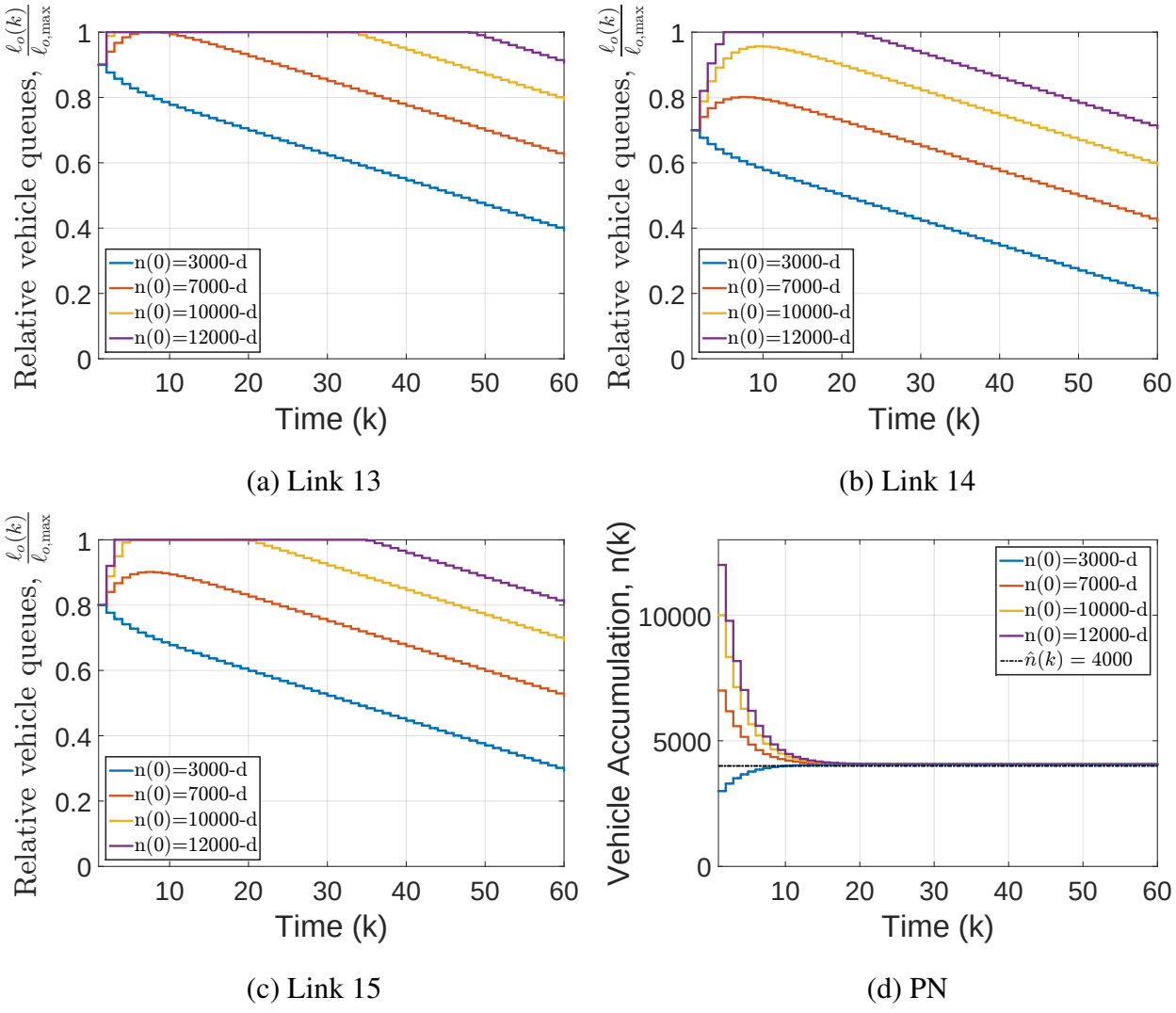


(e) Link 11



(f) Link 12

Figure 4.43: CAP: Relative vehicle queues at links 7–12.



(a) Link 13

(b) Link 14

(c) Link 15

(d) PN

Figure 4.44: CAP: (a)–(c) Relative vehicle queues at links 13–15; (d) State trajectory of PN.

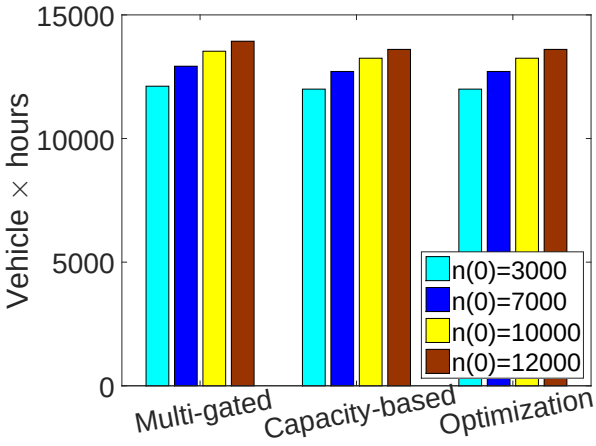
4.5.4 Performance assessment

This section compares the proposed multi-gated perimeter flow control strategy (see Section 4.3), which explicitly considers the queue dynamics outside the protected network area and operational constraints (minimum/maximum queues, capacities, etc.), with the two perimeter-ordered flow allocation policies presented in Section 4.4, namely CAP and OAP. All strategies were first compared for a *scenario without external demand*, designed to allow for CAP/OAP to cope with excessive queues at origin links (outside the protected network area) and effectuate a fair comparison. Then two *scenarios with medium and high external demand* were used to demonstrate the equity properties of MGC and its ability to manage excessive queues outside of the protected network area and optimally distribute the input flows with respect to geometric characteristics.

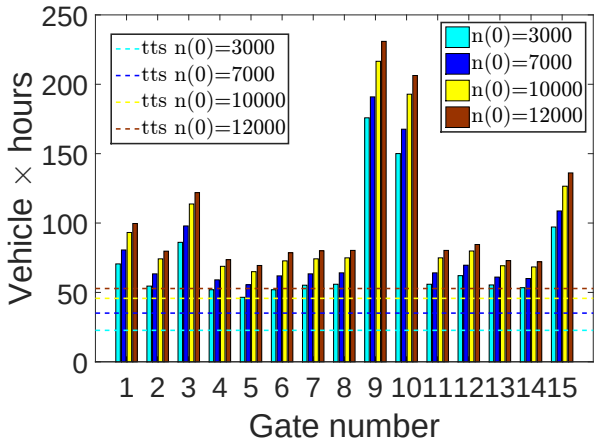
4.5.4.1 Comparison between MGC and CAP/OAP

Figure 4.45 depicts the obtained results (delays/TTS) within the protected network and gated links under MGC, CAP, and OAP for four initial states in a scenario without external demand. Figure 4.45(a) underlines the ability of all strategies to protect the inner network area (PN) of downtown San Francisco from congestion with similar performance across different traffic conditions (ranging from free-flowing to near gridlock). Table 4.6 depicts the average TTS for each strategy including 'no control' scenario across different initial traffic states. Although TTS within the protected network under MGC is slightly higher compared to OAP/CAP, MGC manages to significantly improve the queues and TTS outside the protected network area. In addition, the TTS for No control is the lowest among other strategy, since it process less cars but keeps queues higher outside the protected network. Comparing Figure 4.45(b) with Figures 4.45(c) and 4.45(d), TTS (and delay) under MGC is seen to be four times lower at any gate compared to TTS under OAP/CAP. Table 4.7 further supports the above arguments on the superiority of MGC versus OAP/CAP and No control particularly at the original links outside the protected network. Note that the incoming/gated flows from the periphery of the network is much higher under MGC compared to OAP/CAP. In other words, the MGC strategy serves more cars and maximises throughput for the same simulation horizon shown by the improvement value 93% compared to No control (cf. relative queues at origin links under MGC in Figures 4.23–4.25 (the "-d" trajectories) and OAP/CAP in Figures 4.42–4.44; Figures 4.36–4.38). Clearly under no control, OAP/CAP a high number of trips is unserved at the end of simulation.

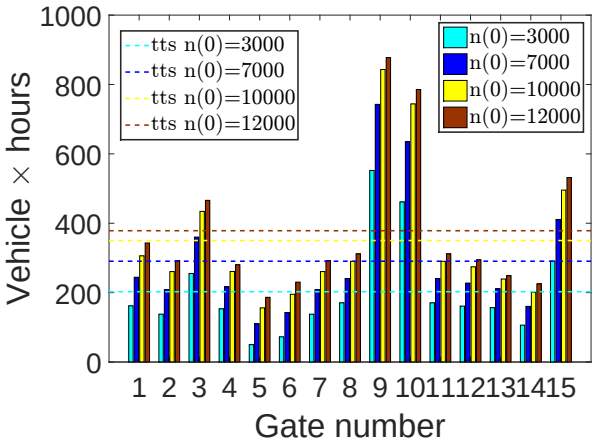
It should be noted that under OAP and CAP drivers experience the same TTS on average, within or outside the protected network area. Though slight differences can be observed from a careful inspection of Figures 4.45(c) and 4.45(d). This can be also confirmed from the relative queues in Figures 4.42–4.44 and Figures 4.36–4.38. This strange result attributed to the lack of flexibility of both allocation strategies to manage the developed queues outside the protected



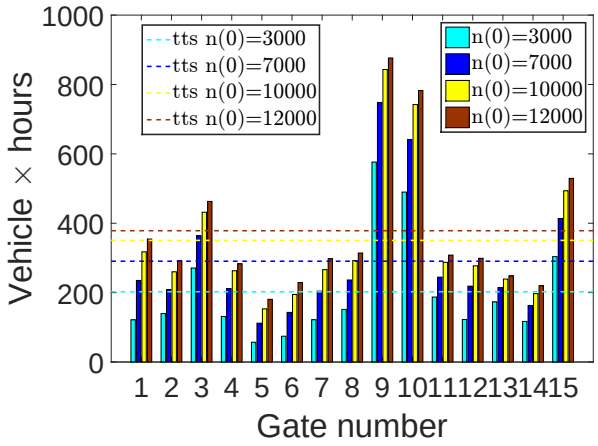
(a) TTS at PN for MGC, CAP, and OAP



(b) TTS at gated links with MGC



(c) TTS at gated links with CAP



(d) TTS at gated links with OAP

Figure 4.45: Total time spent within the protected network and gated links under MGC, CAP, and OAP for four initial states in a scenario without external demand.

Table 4.6: Average value of TTS at PN under No control, MGC, CAP, and OAP for different initial states in a scenario without external demand.

Policy	$n(0) = 3000$ veh	$n(0) = 7000$ veh	$n(0) = 10000$ veh	$n(0) = 12000$ veh	units
No control	11798	12607	13213	13618	veh-h
Improvement	-	-	-	-	%
MGC	12116	12925	13531	13936	veh-h
Improvement	2.7	2.5	2.4	2.3	%
CAP	11998	12712	13248	13605	veh-h
Improvement	1.7	0.83	0.26	-0.1	%
OAP	11998	12712	13248	13605	veh-h
Improvement	1.7	0.83	0.26	-0.1	%

Table 4.7: Average value of TTS at gated links under No control, MGC, CAP, and OAP for different initial states in a scenario without external demand.

Policy	$n(0) = 3000$ veh	$n(0) = 7000$ veh	$n(0) = 10000$ veh	$n(0) = 12000$ veh	units
No control	317	317	317	317	veh-h
Improvement	-	-	-	-	%
MGC	22.67	35.09	45.72	52.81	veh-h
Improvement	93	89	86	83	%
CAP	202.65	290.56	350.16	378.47	veh-h
Improvement	36	8.3	-10.5	-19.4	%
OAP	202.58	290.56	350.59	378.45	veh-h
Improvement	36	8.3	-10.5	-19.4	%

network area. Long queues at origin links (without knowledge of queue dynamics and geometric characteristics) result to conservative control and more or less similar releasing rates.

Table 4.8 displays estimated queues that are unserved at the end of simulation. For MGC, the protected network is empty and queues are dissolved at around $k = 45$. Meanwhile, the other approaches including no control still have queues remaining to dissolve and may take longer time. This observation highlights that although MGC experienced high TTS for PN, it accommodates more traffic and serve more trips in PN while still provides significant improvement in queue lengths (exterior performance). Although it seems that MGC forces more traffic into the PN, it actually consider interference of the created queues at the gated links which may lead to overestimating the benefits of the gating. Traditional gating may induce imbalance gating of vehicles and consequently lead to long queues and overflow phenomena within origin links. Furthermore the control objectives are to equalise the relative vehicle queues as well as maintaining vehicle accumulation in protected network around desired point. The aims are fulfill if the weightage

Table 4.8: Vehicle queues under No control, MGC, CAP, and OAP for different initial states in a scenario without external demand.

Policy	$n(0) = 3000$ veh	$n(0) = 7000$ veh	$n(0) = 10000$ veh	$n(0) = 12000$ veh
No control	1583	1583	1583	1583
MGC	0	0	0	0
CAP	537	998	1348	1582
OAP	532	997	1348	1582

in control framework is selected appropriately which will reduce the risk of overspilling links while protecting the network. Hence the choice of weightage $\mathbf{Q}$ as explained in Section 4.3 is seen as physically reasonable.

4.5.4.2 Equity properties of MGC

Figure 4.46 demonstrates the equity properties of the proposed approach to better manage excessive queues outside of the protected network area and optimally distribute the input flows with respect to geometric characteristics. These figures depict results obtained for two different demand scenarios, namely medium and high, and four initial states $n(0)$ (congested and semi-congested regimes) of the fundamental diagram. As can be seen in Figure 4.46(a), Total Time Spent (TTS) within the protected network area increases with vehicle accumulation for both demand scenarios. Remarkably, Figures 4.46(b)–4.46(c) demonstrate the equity properties of the proposed multi-gated perimeter flow control. More precisely, gates with similar geometric characteristics experience similar TTS (delays) for two different demand scenarios and four different initial states in the protected network area. For instance, three different groups of gates can be distinguished with similar TTS, Group A including gates 2, 4–8, 11–14; Group B including gates 1, 3, 15; and Group C including gates 9, 10. Contrasting gates in Groups A, B, and C with the geometric characteristics in Table 4.1, further supports the equity properties of the proposed multi-gated control.

Tables 4.9 and 4.10 present the average TTS at protected network and gated links, respectively. As can be seen (cf. Table 4.10 (high demand) with Table 4.7 (medium demand)), delays incurred at gated links under MGC for the high demand scenario are comparable or lower (for saturated conditions) to those obtained under CAP/OAP under the medium demand scenario. Note that CAP/OAP were not be able to cope with the high demand scenario (both incurred excessive queues and delays near gridlock), which is attributed to the lack of knowledge of the traffic dynamics outside of the protected network area (no entrance link dynamics).

Concluding, the control flexibility and efficiency of the proposed control (for a number of performance assessment criteria – qualitative and quantitative) while explicitly considering the queue dynamics and constraints underlines the clear superiority of appropriate multi-gated

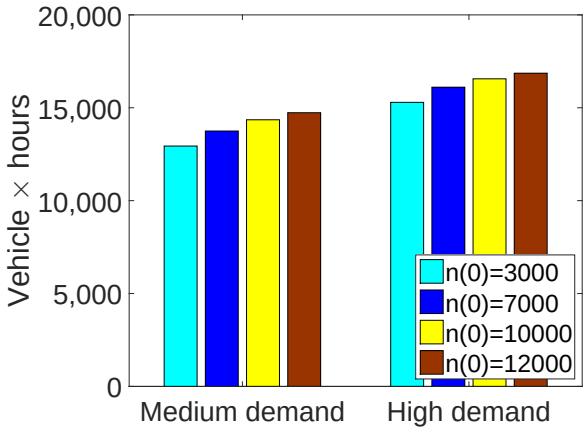
perimeter flow control. Certainly, it should be highlighted the efficiency and equity properties of the proposed multi-gated scheme to better manage excessive queues outside of the protected network area and optimally distribute the input flows.

4.6 Concluding remarks

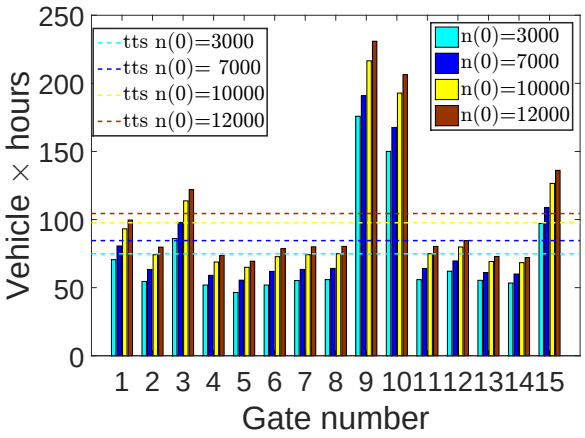
This chapter proposed an integrated model for multi-gated perimeter flow control for an efficient control scheme; by providing additional state variable to describe link dynamics at the outside of the protected area. The proposed scheme determines optimal distribution of input flows across the perimeter of single regions, which allows for the consideration of geometric characteristics at individual gated links, state and control constraints. The perimeter flow control without entrance link dynamics (without constraints) are also examined, which are then applied to two flow allocation policies namely capacity-based flow allocation policy (CAP) and optimisation-based flow allocation policy (OAP).

The developed MGC and allocation policies control schemes were analysed under different case studies (e.g. with or without demand) and applied to the tested area of downtown San Francisco with fifteen gates of different geometric characteristics. Simulation results of MGC demonstrated the optimal flow distribution facilitate a balance relative vehicle queues at links, maintain the vehicle accumulation in the protected network around a desired point, while the system's throughput is maximised. Thus, it exhibits an efficiency and equity properties of the MGC approach to better manage excessive queues outside of the protected network area.

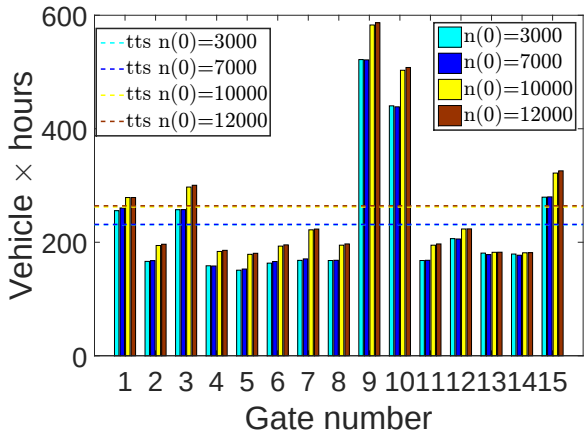
The flow allocation policies, CAP and OAP indicated more or less the same performance for the considered simulation scenarios, where both demonstrated slow performance in dissolving the developed queues at the origin links. The TTS performance for MGC observed to be profoundly lower compared to TTS of both flow allocation policies. This underlined that MGC strategy serves more cars and maximises throughput for the same simulation horizon. By considering the queue dynamics and constraints underlines the clear superiority of appropriate MGC scheme. The equity properties of perimeter flow control have not attracted considerable attention in the literature, although it is an important characteristic of any practical perimeter flow control application.



(a) TTS at PN for MGC



(b) Gated links medium demand



(c) Gated links high demand

Figure 4.46: Total time spent within the protected network area and gated links under MGC for different initial states and two scenarios with external demand.

Table 4.9: Average value of TTS at PN under MGC for different initial states and two scenarios with external demand.

Type of demand	n(0) = 3000 veh	n(0) = 7000 veh	n(0) = 10000 veh	n(0) = 12000 veh
Medium	12939	13748	14354	14735
High	15292	16108	16559	16861

Table 4.10: Average value of TTS at gated links under MGC for different initial states and two scenarios with external demand.

Type of demand	n(0) = 3000 veh	n(0) = 7000 veh	n(0) = 10000 veh	n(0) = 12000 veh
Medium	74.83	84.53	97.67	104.42
High	230.93	231.4	262.18	264.22

Chapter 5

Conclusions and future work

This chapter summarises the two studied problems, their results, and discusses how the thesis contributes to the state-of-the-art. Some suggestions and recommendations for further research conclude the thesis.

5.1 Overview

This thesis has investigated how the so-called macroscopic or network fundamental diagram can be used for the monitoring and control of transport networks. It offers potential sensor selection methods to construct an operational or sparse-measurement network-wide fundamental diagrams. Furthermore, it also proposes a multi-gated perimeter flow control strategy that explicitly considers the queue dynamics and constraints outside of the protected network area for congestion mitigation in monocentric cities. Two case studies, benefited from data-driven inference and several simulation experiments, have shown the effectiveness of the proposed information-theoretic framework for sensor selection and rolling-horizon approach for multi-gated perimeter flow control.

The thesis tackled the following two problems:

Traffic monitoring problem: One can utilise the macroscopic fundamental diagram of network traffic to design perimeter control strategies in order to tackle congestion in monocentric or heterogenous cities with multiple pockets of congestion. To this end, one needs to construct the NFD by flow-occupancy (or density) measurements of a number of inductive-loop detectors (or other sensors) placed at appropriate network locations. A sparse-measurement diagram is constructed by selecting only a smaller number of sensors. Clearly, different levels of network coverage in terms of selected sensors (virtually percentage of monitored links) can provide different levels of accuracy. In this context, installing sensors everywhere in the entire network is costly. However using brute-force search or exhaustive search to reduce the number of sensors and the corresponding measurements would be overwhelming and practical impossible. Thus efficient algorithms for

sensor selection is important for the development of operational or sparse-measurement NFDs, which involve less sensors and corresponding measurements. Chapter 3 offered, for the first time in the relevant literature, an information-theoretic framework for the sensor selection problem to construct sparse-measurement NFDs which preserve the main properties of the full-measurements NFD.

Perimeter flow control problem The ultimate goal of perimeter flow control is to maintain the vehicle accumulation (or proxies of accumulation) in a protected network area around a desired (or critical) point, while the network's throughput is maximised. To achieve this, a perimeter flow control policy "meters" the input flow to the system and hold vehicles outside of the protected network area if necessary, so as to maximise the network throughput. This is usually achieved by controlling with traffic lights, toll stations or automatic billing systems located at the periphery, the perimeter of a protected network area, which exhibits a well-defined NFD. However, gated queues at origin links must be restricted to avoid interference with adjacent street traffic outside of the protected network area and geometric characteristics of the different gates must be taken into account in the optimisation. Chapter 4 addressed this problem and offered efficient control schemes for the optimal distribution of input flows across the perimeter of single regions or the boundaries of neighbour regions which allows for the consideration of geometric characteristics at individual gated links, and operational constraints. Thus, limited origin links storage capacity/geometric characteristics and the requirement of equity for drivers using different gates to enter a protected area are the main reasons towards multi-gated perimeter flow control, in which the thesis contributes to the state-of-the art.

5.2 Discussion and conclusions

Overall, this thesis developed advanced tools and traffic management strategies to mitigate traffic congestion in monocentric cities. Several works in the past have shown that the NFD can be utilised to design efficient perimeter flow control strategies. This design requires an accurate estimation of the NFD from observations from loop detectors (or other sensors) located at appropriate locations in the network. Some of the highlights based on the achieved results are discussed in the sequel.

For the traffic monitoring problem, there is lack in the relevant literature of an automated or data-driven framework for efficiently building sparse-measurements NFDs. The state-of-the-art relies solely on ad-hoc strategies (e.g. random selection or based on brute-force/computer simulation). Besides the existing literature mainly addresses the minimum required level of network coverage while the selection of sensors (or monitored locations) in a network are under study. These limitations from previous works are considered and a generic information-theoretic framework for the optimal selection of traffic sensors is developed to accommodate the construc-

tion of an operational NFD. Such approach is not only helpful in improving the estimation of NFD but also in reducing installation and maintenance costs associated with minimum number of sensors.

The sensor selection problem is in general a combinatorial optimisation problem (NP-hard) involving noisy measurements. To tackle the stochasticity of spatiotemporal phenomena in traffic dynamics, a data-driven GIP model is developed where three different data sources (flows, occupancies, and spatial distance) is used to achieve a good approximation of an operational NFD. The data-driven approach helped to develop a "measure of distance" based on the correlation of occupancies and/or correlation of flows. Additional information such as the spatial distance was used for refining our approach. A number of constraints were then generated and incorporated in the GIP model. Several tools to assess the GIP solutions namely cost of the GIP solution, KL-divergence, and RMSE was used.

Results obtained from the GIP-based rival models were very encouraging. Yet there were many solutions providing more or less the same performance (in terms of KL-divergence) and it was challenging to improve one solution to another. This could be attributed to the size and topology of network used in this study. The availability of real empirical data for this network though gave us some good evidence on the approximation problem and performance of different models. It is expected that a larger heterogeneous network with multiple centers of congestion would make the performance of blind or other ad-hoc strategies to deteriorate.

The rival models (Model 1, 2, 3, 4) were better at maintaining the nominal NFD characteristics such as critical occupancy and flow capacity value although the respective KL-divergence was not necessarily low. In particular, Model 3 used flow and occupancy information provided lower RMSE but demonstrated high KL-divergence. Meanwhile, Model 4 obtained with all the criteria showed an improvement in KL-divergence, though high in RMSE. These observations were in line with our aim to achieve a good approximation and KL-divergence, the model should include more criteria.

The blind strategy indicated a good fit though the capacity flow exceeded the acceptable error. Another drawback of this ad-hoc method is that it is totally based on random link selection. For a large network with multiple pockets of congestion, the disparity of occupancies and flows will affect the performance of blind or other ad-hoc strategies. The radius strategy could be considered as alternative but its performance was not consistent across different radius/areas. In other words, it is location-sensitive or constrained. This confirmed our initial conjecture and motivation to pursue this work, that the spatial distance cannot be employed as a single metric but can be combined with other sources of information (as in the developed GIP-based models). It is also observed regardless of the model or strategy used, when the correlation increased, the cost of GIP also increased (property of the model). This included more links in the solutions contributed to the information gain therefore KL-divergence decreased (showing improvement). The evidence from this study suggest that: (a) using correlation between sensors in a network

to quantify "redundancy" of sensor selections is a very promising tool, opposing to "spatial distance" in conventional methods; and, (b) KL-divergence is more generic and reliable metric of information loss, compared to existing methods in the relevant literature e.g., to use density ratios.

One limitation of our data-driven inference scheme is that information/measurements from pre-installed detectors over the entire network need to be collected and checked beforehand. In other words, the entire or part of the network needs to be sensed first before one can apply the proposed measure of correlation between random variables as a "distance metric". Nevertheless, the advantage is that there is no formal linearity assumption regarding the measurement model and its distribution (e.g. Gaussian) needed as for example in [61]. Instead, the analysed models in this study indicated a geometric distribution, which can be used for inference when a network has incomplete data. Another limitation could be attributed from the difficulty of requiring data for the GIS method applied, since most cities do not have prior traffic information; for example a city that does not have any sensors. With no data availability, GIS method could still be applied for such networks by using information only from Radius as a "distance metric" (see constraint (3.6) or (3.7)). Certainly the obtained operational NFDs can be used for perimeter flow control. This thesis offers some guidance to researchers in traffic engineering and control who are interested in conducting sensor selection problems with an information-theoretic framework in the field of transport networks or other field that involve monitoring of spatiotemporal phenomena.

In the second part of the thesis, most studies assumed that a single ordered input flow obtained from a perimeter control strategy is equally (or not optimally) distributed to a number of candidate junctions (or gated links) at the periphery of the network. The distribution of allowable flow from the perimeter has not been fully considered especially with a scheme that allow for direct considerations of state and control constraints. Moreover, suboptimal methodology of flow distribution to the gated links creates imbalance entrance for drivers to enter the protected network area using different gates. These limitations from previous work was considered and a multi-gated perimeter flow control, an optimisation-based scheme that allows for the consideration of different types of state and control constraints was considered to facilitate congestion control in urban networks.

The improvement of perimeter flow modelling was suggested by entrance link dynamics to account queues at gated links. The suggested controller considered dynamics in both inside and outside protected network area. Simulation results demonstrated the multi-gated perimeter flow control dissolved the queues in a balanced way, depicted the desired control objective of queue balancing and equity for drivers using different gates to enter the protected area was achieved. The suggested controller also stabilised all input flow to their desired input flow value and clearly demonstrated different trajectories at different gated links. On the other hand, the alternative distribution (capacity- and optimisation- based) without consideration of different geometric characteristics of gated links were ineffective to achieve nominal flow and consequently

developed queue lengths. The proposed approach was shown to perform well under demand provision (medium or high). Performance results demonstrated the multi-gated perimeter flow control was capable of improving TTS at gated links while preserving the performance of interior network, which hardly see in the alternative strategies. Hence the multi-gated perimeter flow control demonstrated its efficiency and equity properties to better managed excessive queues and avoid delays outside the protected area while optimally distribute the input flows. Multi-gated perimeter flow control is of great importance for optimising network capacity because it maximises the number of trips completed.

5.3 Future work

Approximating the NFD with lower number of measurements is challenging because one can employ different models to approximate it. In principle, the sensor selection problem is a combinatorial optimisation problem, thus hard-to-solve (known as NP-hard). Sensor measurements stochasticity imposes high uncertainty and there is an extremely high range of "optimal solutions" providing similar or comparable approximations (depending on the method of assessment). Potential approaches to tackle this problem include methods from artificial intelligent/machine learning [5], network observability [85], and Bayesian statistics [11].

For the multi-gated perimeter flow control problem, the equity properties of perimeter flow control have not attracted considerable attention in the literature, although it is an important characteristic of any practical perimeter flow control application. A multi-criteria optimisation approach can be also used to handle different objectives [80]. Future research should focus on: (a) extending the proposed framework for multi-region cities described by three-dimensional bi-modal passenger and vehicle NFDs [9, 16, 38]; (b) investigating the efficiency versus equity properties of perimeter flow control with queue dynamics, and; (c) integrating of additional components for dynamic routing and road pricing [70, 87]. The proposed model could be simulated with full network simulators (e.g. AIMSUM or VISSIM) at a more advanced stage in order to derive more reliable simulations and conclusions pertaining to the comparative method efficiencies.

Appendix A

Information theory

This appendix provides the necessary background on information theory to develop a measure of information correlation, which is used as a "distance" metric to provide sufficient coverage and information accuracy in a set cover problem in Chapter 3. It also presents the Kullback-Leibler divergence (relative entropy) which will be used to measure the dissimilarity between probability mass functions corresponding to different solutions of the developed set cover integer programming problem in Chapter 3. A measure of information correlation is then based on mutual information and joint entropy. The interested reader is referred to [20] for a detailed treatment in information theory.

A.1 Entropy

Let X be a discrete random variable that is completely defined in a finite set $\mathcal{X} = \{0, 1, 2, \ldots\}$. The value $p_X(x) = \mathbf{P}(X = x)$ is the probability that the random variable X takes the value x. Then $p(x)$ defines a probability mass function (pmf) for the discrete random variable X with support $\mathcal{X}$. In the sequel, all random variables and distributions are considered discrete.

The (Shannon) entropy of a random variable X or its probability mass function $p_X(x)$ is a non-negative measure of the amount of "uncertainty" in the distribution [78].

Definition 2 (Entropy)**.** *The entropy $H[p(x)]$ of a distribution $p(x)$ is defined by*

$$H[p(x)] \triangleq -\sum_{x \in \mathcal{X}} p(x) \log p(x) = -\mathbf{E}[\log p(X)]\,, \tag{A.1}$$

when the sum exists.

The operator $\mathbf{E}_p[\cdot]$ denotes expectation, where X is drawn according to the distribution $p(x)$. Entropy is measured in "bits" if the $\log$ is to the base 2 and in "nats" (natural units of information) in the case of the natural (to the base e) $\log$. Here base two logarithms are used and $0 \log 0$ is defined to be 0.

A.2 Relative entropy and measure of information correlation

The relative entropy or the Kullback-Leibler divergence (KL-divergence in short) of two distributions is denoted as $\Delta(p \,\|\, q)$. It is a measure of the inefficiency of assuming that the distribution is q when the true (reference) distribution is p. It arises as an expected logarithm of the likelihood ratio [63].

Definition 3 (Relative Entropy; Kullback-Leibler Divergence)**.** *Let $p(x)$, $q(x)$, $x \in \mathcal{X}$, be two probability mass functions. The relative entropy of $q(x)$ with respect to $p(x)$ (reference distribution), or the Kullback-Leibler divergence of $q(x)$ from $p(x)$ is defined as*

$$\Delta(p \,\|\, q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} = \mathbf{E}_p \left[\log \frac{p(X)}{q(X)} \right]. \tag{A.2}$$

Here $0 \log 0/p$ and $0 \log 0/0$ are defined to be 0 while $p \log p/0$ is defined to be ∞. In general the KL-divergence is not symmetric under interchange of the distributions p and q (i.e., $\Delta(p \,\|\, q) \neq \Delta(q \,\|\, p)$), and it does not obey the triangle inequality (i.e. it is not a distance measure). Nonetheless, it is enough like a metric that it can be used to construct a geometry on the space of probability mass functions, and so of statistical models.

Proposition 1 (Jensen's inequality)**.** *Let f be convex (respectively, strictly convex). Then for any random variable X, $\mathbf{E}[f(X)] \geq (>)\, f(\mathbf{E}[X])$, with equality if X is constant.*

Proof. Omitted, for a proof see e.g., [20]. □

Convergence of probability distributions, $p^n \to p$, means point-wise convergence, that is, $p^n(x) \to p(x)$ for each $x \in \mathcal{X}$. A key property of KL-divergence is that it is nonnegative and zero if and only if two distributions are equal.

In general the KL-divergence is unbounded from above, since one can find distributions that are arbitrarily close in total variation but with arbitrarily high relative entropy. The following lemma, known as Pinsker's inequality, gives a lower bound on the relative entropy in terms of the total variation distance.

Lemma 1 (Pinsker's inequality: Divergence and Total Variation)**.** *For any two probability mass functions $p(x)$, $q(x)$, $x \in \mathcal{X}$, $\Delta(p \,\|\, q) \geq \frac{1}{2 \ln 2} \|p - q\|_1^2$.*

Proof. Omitted, for a proof see e.g., [20]. □

This lemma suggests that convergence in relative entropy, $\Delta(p \,\|\, q^n) \to 0$ as $n \to \infty$, where q^n is a sequence of rival distributions, implies convergence in the total variation ℓ_1 metric.

Consider two random variables X defined in a finite set $\mathcal{X}$ and Y defined in a finite set $\mathcal{Y}$ with marginal probability mass functions $p(x)$ and $p(y)$, respectively; a joint probability mass function $p(x, y)$ and a conditional probability mass function $p(x \,|\, y)$. Similarly to the definition of the entropy of a single variable, the joint entropy of a pair of random variables is defined.

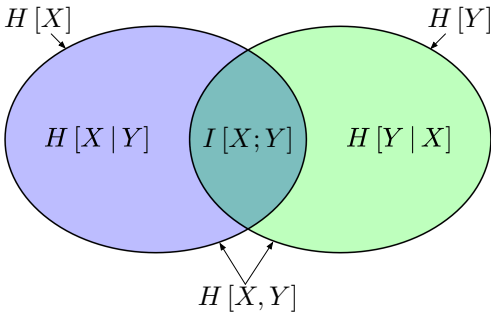


Figure A.1: Venn diagram showing the relationship between entropy (conditional, joint) and mutual information.

Definition 4 (Joint Entropy)**.** *The joint entropy $H[X,Y]$ of two discrete variables X and Y with a joint probability mass function $p(x,y)$ is defined as*

$$H[X,Y] \triangleq -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y) = -\mathbf{E}[\log p(X,Y)]\,. \tag{A.3}$$

The joint entropy represents the amount of information needed on average to determine the value of two discrete random variables.

Definition 5 (Conditional Entropy)**.** *The conditional entropy $H[X\,|\,Y]$ of X given Y is defined as*

$$H[X\,|\,Y] \triangleq -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x\,|\,y)\,. \tag{A.4}$$

Mutual information is the reduction in uncertainty of a random variable conditional on the knowledge of another random variable.

Definition 6 (Mutual Information)**.** *The mutual information between two random variables, X and Y, is the divergence of the product of their marginal distributions from their actual joint distribution:*

$$\begin{aligned} I\,[X;Y] &\triangleq \Delta(p(x,y)\,\|\,p(x)\,p(y)) \\ &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)\,p(y)} = \mathbf{E}_{p(x,y)}\left[\log\frac{p(X,Y)}{p(X)\,p(Y)}\right]. \end{aligned} \tag{A.5}$$

The mutual information $I\,[X;Y]$ is symmetric in X and Y and always non-negative and is equal to zero if and only if X and Y are independent.

The following definition provides a measure of information correlation between two random variables with the use of the conditional, joint entropy and mutual information.

Definition 7 (A measure of information correlation)**.** *The quantity*

$$\varrho(X,Y) = \frac{I\,[X;Y]}{H[X,Y]}, \tag{A.6}$$

is a measure of information correlation between the random variables X and Y, where $0 \leq \varrho \leq 1$. Furthermore, if X and Y are identically distributed, but not necessarily independent then the measure of information correlation is given by

$$\varrho(X,Y) = \frac{I\left[X;Y\right]}{H[X]}. \tag{A.7}$$

The measure of information correlation ϱ given by (A.6) or (A.7) is zero if and only if X and Y are independent, while ϱ is one if and only if X and Y have a one-to-one relationship.

Another measure of linear correlation between two (continuous) random variables is given by the classic Pearson correlation coefficient

$$\rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X\,\sigma_Y} = \frac{\mathbf{E}[(X-\mathbf{E}[X])\,(Y-\mathbf{E}[Y])]}{\sigma_X\,\sigma_Y}, \tag{A.8}$$

where $\operatorname{cov}(X,Y)$ is the covariance of X and Y while σ_X and σ_Y are the standard deviations of X and Y, respectively. The correlation coefficient is bounded by $-1 \leq \rho \leq 1$ and it is symmetric. Two random variables X and Y are uncorrelated when their correlation coefficient is zero. They are perfectly correlated if $\rho = 1$ and anti-correlated if $\rho = -1$. Consider a random vector $\mathbf{X}$ of random variables $X_1, X_2, \ldots, X_n$. The correlation matrix of $\mathbf{X}$ is defined as $\operatorname{corr}(\mathbf{X}) = [\rho(X_i, X_j)]_{i,j=1,\ldots,n}$, where $\operatorname{diag}(\operatorname{corr}(\mathbf{X})) = \mathbf{1}_n$; and, it is a symmetric matrix.

Appendix B

Model predictive control

This appendix provides the necessary background on the linearisation of nonlinear systems and introduces model predictive control (MPC) for the constrained control of linear discrete-time systems. MPC is particularly suitable to control problems involving hard state and control constraints and, thus it is employed for the solution of the multi-gated perimeter flow control developed in Chapter 4.

B.1 Linearisation of nonlinear systems

A dynamical system with nonlinearities can be approximated by a linear model so that one can design and apply simple linear control. Consider the nonlinear system described by,

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), \mathbf{d}(k)], \tag{B.1}$$

$$\mathbf{y}(k) = \mathbf{g}[\mathbf{x}(k)], \tag{B.2}$$

where $\mathbf{x} \in \mathbb{R}^n, \mathbf{u} \in \mathbb{R}^m, \mathbf{d} \in \mathbb{R}^q, \mathbf{y} \in \mathbb{R}^p$ are the state, control, disturbance and output vectors respectively. $\mathbf{f} \in \mathbb{R}^n$ describe the dynamic process and $\mathbf{g} \in \mathbb{R}^p$ describe the dynamic output where both are nonlinear vector functions. Given a desired steady-state $(\hat{\mathbf{x}}, \hat{\mathbf{u}}, \hat{\mathbf{d}}, \hat{\mathbf{y}})$, denote $\Delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}$, $\Delta\mathbf{u} = \mathbf{u} - \hat{\mathbf{u}}$, $\Delta\mathbf{d} = \mathbf{d} - \hat{\mathbf{d}}$ and $\Delta\mathbf{y} = \mathbf{y} - \hat{\mathbf{y}}$. Linearisation of (B.1) and (B.2) around the desired steady-state yields,

$$\Delta\mathbf{x}(k+1) = \mathbf{A}\Delta\mathbf{x}(k) + \mathbf{B}\Delta\mathbf{u}(k) + \mathbf{H}\Delta\mathbf{d}(k) \tag{B.3}$$

$$\Delta\mathbf{y}(k) = \mathbf{Y}\Delta\mathbf{x}(k) \tag{B.4}$$

where $\Delta\mathbf{x}, \Delta\mathbf{u}, \Delta\mathbf{d}$ and $\Delta\mathbf{y}$ are the linearised state, control, disturbance and output vectors. $\mathbf{A} = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_{\hat{\mathbf{x}}}$, $\mathbf{B} = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\right]_{\hat{\mathbf{u}}}$, $\mathbf{H} = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{d}}\right]_{\hat{\mathbf{d}}}$ and $\mathbf{Y} = \left[\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\right]_{\hat{\mathbf{x}}}$ are the state, control, disturbance and output matrices, respectively.

B.2 Model predictive control

B.2.1 Basic components of MPC

Model predictive control (MPC) is a model-based control procedure where the goal is to include physical constraints (state and control constraints) in the controller design phase and then applied for the solution of optimal control problems [68]. The three basic components of MPC are prediction, optimisation, and rolling horizon, and briefly described in the sequel [15].

Prediction

Predictive control uses predictions of the system's evolution over a finite horizon and making decisions on the optimal control strategy. In the case of a discrete-time linear system, the state space representation is given by,

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k), \tag{B.5}$$

where $\mathbf{x}(k)$ and $\mathbf{u}(k)$ are the state and control variables at time instant k. $\mathbf{A}$ and $\mathbf{B}$ denote the state and control matrices, respectively. If (B.5) is extended for a prediction horizon of length N, an augmented state space model can be obtained for the MPC. The predicted sequence of controls and states is stacked into the vectors $\mathbf{U}$ and $\mathbf{X}$ as follows,

$$\mathbf{U}(k) = \begin{bmatrix} \mathbf{u}(k) \\ \mathbf{u}(k+1) \\ \vdots \\ \mathbf{u}(k+N-1) \end{bmatrix}, \quad \mathbf{X}(k) = \begin{bmatrix} \mathbf{x}(k+1) \\ \mathbf{x}(k+2) \\ \vdots \\ \mathbf{x}(k+N) \end{bmatrix}. \tag{B.6}$$

Here $\mathbf{u}(k+i)$ and $\mathbf{x}(k+i)$ denote the control and state vectors at the predicted time-step $k+i$ with index $i = 1, 2, \ldots, N$, at current time k. At time step $k+0$ ($i = 0$), the initial condition is defined as $\mathbf{x}(k)$. Therefore $\mathbf{x}(k+i)$ evolves according to the prediction model:

$$\mathbf{x}(k+i+1) = \mathbf{A}\mathbf{x}(k+i) + \mathbf{B}\mathbf{u}(k+i), \quad i = 0, 1, \ldots \tag{B.7}$$

Specifically, the predicted state sequence generated by the linear state space model (B.7) with input sequence $\mathbf{U}(k)$ can be written as,

$$\begin{aligned}
&\mathbf{x}(k) \\
\mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) \\
\mathbf{x}(k+2) &= \mathbf{A}^2\mathbf{x}(k) + \mathbf{A}\mathbf{B}\mathbf{u}(k) + \mathbf{B}\mathbf{u}(k+1) \\
&\vdots
\end{aligned}$$

Optimisation

The predictive control input is computed by minimising a cost function (criterion), which has predicted sequences $\mathbf{u}$ and $\mathbf{x}$. The cost function can take various forms, but generally a quadratic cost is applied in MPC over the prediction horizon N:

$$J(k) = \sum_{k=0}^{N-1} \Big(\|\mathbf{x}(k+i)\|^2_{\mathbf{Q}} + \|\mathbf{u}(k+i)\|^2_{\mathbf{R}} \Big), \tag{B.8}$$

where $\mathbf{Q}$, $\mathbf{R}$ are positive definite weighting matrices. For a given initial state $\mathbf{x}_0$, $J(k)$ is a function of $\mathbf{u}(k)$ and the optimal control input is computed by minimising $J(k)$ subject to the state and control constraints:

$$\min_{\mathbf{u}} J(k) \tag{B.9}$$

$$\text{subject to} \quad \mathbf{u}(k+i) \in \mathbb{U}, \tag{B.10}$$

$$\mathbf{x}(k+i) \in \mathbb{X}, \tag{B.11}$$

where $\mathbb{U}$ and $\mathbb{X}$ denote the polyhedral constraint sets of inputs and states, respectively. The resulting optimal control input is denoted as $\mathbf{u}^*(k)$.

Rolling horizon

At the end of the optimisation process, only the first element of the optimal predicted sequence $\mathbf{u}^*(k)$ is applied to the process (see the top diagram in Figure B.1). At the next time step $k+1$ a new optimal control problem based on new measurements of the state is solved over a shifted horizon (see bottom diagram in Figure B.1). The process of computing $\mathbf{u}^*(k)$ through minimising the predicted cost and implementing the first element of $\mathbf{u}^*$ is then repeated at each sampling instant $k = 0, 1, \ldots$.

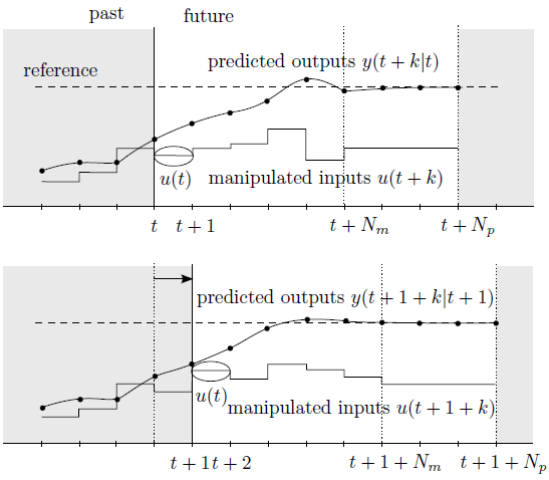


Figure B.1: The rolling horizon idea behind MPC [12].

B.2.2 Unconstrained MPC

Consider a linear time-invariant system described by the following state-space model [40]:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k), \tag{B.12}$$

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{d}(k), \tag{B.13}$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ is the state, $\mathbf{u}(k) \in \mathbb{R}^m$ is the control input, $\mathbf{y}(k) \in \mathbb{R}^m$ is the output and $\mathbf{d}(k) \in \mathbb{R}^m$ is disturbances at k-th sampling instant. Denote the set point or desired steady-state values for $\mathbf{x}(k)$, $\mathbf{u}(k)$ and $\mathbf{d}(k)$ as $\hat{\mathbf{x}}$, $\hat{\mathbf{u}}$ and $\hat{\mathbf{d}}$; $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{m\times m}$ are state and control matrices respectively. Denote $\Delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}$, $\Delta\mathbf{u} = \mathbf{u} - \hat{\mathbf{u}}$ and $\Delta\mathbf{d} = \mathbf{d} - \hat{\mathbf{d}}$.

Consider the problem of tracking a constant reference signal $\hat{\mathbf{y}}$ to achieve zero output error $\mathbf{y}(k)$ in (B.13), that is,

$$\begin{aligned} \mathbf{e}(k) &= \mathbf{y}(k) - \hat{\mathbf{y}}, \\ \mathbf{e}(k) &= (\mathbf{C}\mathbf{x}(k) + \mathbf{d}(k)) - (\mathbf{C}\hat{\mathbf{x}}(k)) + \hat{\mathbf{d}}(k)), \end{aligned} \tag{B.14}$$

Similarly,

$$\Delta\mathbf{e}(k) = \mathbf{C}\Delta\mathbf{x}(k) + \Delta\mathbf{d}(k). \tag{B.15}$$

As in the previous section our goal is to minimise the quadratic objective function,

$$J(k) = \frac{1}{2}\sum_{k=0}^{N_o-1}\Big(\|\Delta\mathbf{e}(k)\|^2_{\mathbf{Q}} + \|\Delta\mathbf{u}(k)\|^2_{\mathbf{R}}\Big), \tag{B.16}$$

where $\mathbf{Q}$, $\mathbf{R}$ are positive semi-definite and positive definite diagonal weighting matrices, respectively. N_o is the prediction horizon and let N_p to be the control horizon.

The minimisation problem is solved at each time step for the current state and disturbances values. Then, the first control move is applied to the plant and the procedure is carried out again. This rolling-horizon procedure closes the loop that is avoids myopic control actions while embedding a dynamic open-loop optimisation problem in a responsive environment. Predicted demand flows $\mathbf{d}(k)$ may be calculated by use of historical information or suitable extrapolation methods.

Given the known initial state $\mathbf{x}(0) = \mathbf{x}_0$, a static convex optimisation problem may be formulated over N_o due to the discrete-time nature of the involved process. The following equations describe the evolution of the system:

$$\begin{aligned}
\Delta\mathbf{x}(1) &= \mathbf{A}\Delta\mathbf{x}(0) + \mathbf{B}\Delta\mathbf{u}(0), \\
\Delta\mathbf{x}(2) &= \mathbf{A}^2\Delta\mathbf{x}(0) + \mathbf{AB}\Delta\mathbf{u}(0) + \mathbf{B}\Delta\mathbf{u}(1), \\
&\vdots \\
\Delta\mathbf{x}(N_o) &= \mathbf{A}^{N_o}\Delta\mathbf{x}(0) + \mathbf{A}^{N_o-1}\mathbf{B}\Delta\mathbf{u}(0) + \ldots + \mathbf{B}\Delta\mathbf{u}(N_p-1),
\end{aligned}$$

or in tabular form:

$$\begin{bmatrix} \Delta\mathbf{x}(k+1) \\ \Delta\mathbf{x}(k+2) \\ \vdots \\ \Delta\mathbf{x}(k+N_o) \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^{N_o} \end{bmatrix} \Delta\mathbf{x}(0) + \begin{bmatrix} \mathbf{B} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{AB} & \mathbf{B} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \\ \mathbf{A}^{N_o}\mathbf{B} & \mathbf{A}^{N_o-1}\mathbf{B} & \ldots & \mathbf{B} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{u}(k) \\ \Delta\mathbf{u}(k+1) \\ \vdots \\ \Delta\mathbf{u}(k+N_p-1) \end{bmatrix}$$

In compact notation:

$$\Delta\mathbf{x}(k) = \mathbf{\Omega}\Delta\mathbf{x}(0) + \mathbf{\Gamma}\Delta\mathbf{u}(k), \quad k = 0, 1, \ldots N_o - 1, \tag{B.17}$$

where,

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{A} \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^{N_o} \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} \mathbf{B} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{AB} & \mathbf{B} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \\ \mathbf{A}^{N_o}\mathbf{B} & \mathbf{A}^{N_o-1}\mathbf{B} & \ldots & \mathbf{B} \end{bmatrix}. \tag{B.18}$$

The state, control and disturbance vectors are defined by,

$$\Delta\mathbf{X} = \begin{bmatrix} \Delta\mathbf{x}(1)^\mathsf{T} & \Delta\mathbf{x}(2)^\mathsf{T} & \cdots & \Delta\mathbf{x}(N_o)^\mathsf{T} \end{bmatrix}^\mathsf{T},$$
$$\Delta\mathbf{U} = \begin{bmatrix} \Delta\mathbf{u}(0)^\mathsf{T} & \Delta\mathbf{u}(1)^\mathsf{T} & \cdots & \Delta\mathbf{u}(N_o-1)^\mathsf{T} \end{bmatrix}^\mathsf{T},$$
$$\Delta\mathbf{D} = \begin{bmatrix} \Delta\mathbf{d}(0)^\mathsf{T} & \Delta\mathbf{d}(1)^\mathsf{T} & \cdots & \Delta\mathbf{d}(N_p-1)^\mathsf{T} \end{bmatrix},^\mathsf{T}$$

and the matrices

$$\begin{aligned} \mathcal{Q} &= \text{diag}\{\mathbf{C}^\mathsf{T}\mathbf{Q}\mathbf{C}, \ldots, \mathbf{C}^\mathsf{T}\mathbf{Q}\mathbf{C}\}, \\ \mathcal{R} &= \text{diag}\{\mathbf{R}, \ldots, \mathbf{R}\}, \\ \mathcal{Z} &= \text{diag}\{\mathbf{C}^\mathsf{T}\mathbf{Q}, \ldots, \mathbf{C}^\mathsf{T}\mathbf{Q}\}. \end{aligned} \tag{B.19}$$

Substitute (B.17) into (B.15) and then into (B.16) and collecting terms to yield:

$$\begin{aligned} J = \frac{1}{2}\Big(&\Delta\mathbf{u}^\mathsf{T}(\mathbf{\Gamma}^\mathsf{T}\mathcal{Q}\mathbf{\Gamma} + \mathcal{R})\Delta\mathbf{u} + 2\Delta\mathbf{u}^\mathsf{T}(\mathbf{\Gamma}^\mathsf{T}\mathcal{Q}\mathbf{\Omega})\Delta x(0) + 2\Delta\mathbf{u}^\mathsf{T}(\mathbf{\Gamma}^\mathsf{T}\mathcal{Z})\Delta\mathbf{d} + \\ &\Delta x(0)^\mathsf{T}\mathbf{\Gamma}^\mathsf{T}\mathcal{Z}\Delta\mathbf{d} + \Delta\mathbf{d}^\mathsf{T}\mathcal{Q}\Gamma\Delta x(0) + \Delta\mathbf{d}^\mathsf{T}\mathcal{Q}\Delta\mathbf{d}\Big) \end{aligned} \tag{B.20}$$

Given that $\mathbf{R} \succ \mathbf{0}$ in the cost criterion (B.16) the Hessian $\mathbf{H}$ is positive definite, and thus the Quadratic Programming (QP) is convex and has a global optimum. In the unconstrained problem, considering the gradient of J with respect to $\mathbf{u}$ and $\text{grad}_\mathbf{u} J = 0$ leads to analytical solution:

$$\Delta\mathbf{U} = -\mathbf{H}^{-1}\mathbf{F}[\mathbf{x}(0) + \mathbf{G}\Delta\mathbf{D}], \tag{B.21}$$

where $\mathbf{H} = \mathbf{\Gamma}^\mathsf{T}\mathcal{Q}\mathbf{\Gamma} + \mathcal{R}$ is the Hessian of the corresponding quadratic program (QP), $\mathbf{F} = \mathbf{\Gamma}^\mathsf{T}\mathcal{Q}\mathbf{\Omega}$, and $\mathbf{G} = \mathbf{\Gamma}^\mathsf{T}\mathcal{Z}$.

B.2.3 Constrained MPC

In the constrained case of MPC, however, there does not exist any analytic solution. Instead, the idea in MPC is to define a prediction horizon N_p and approximate the problem with a finite horizon cost. Inequality constraints are introduced into the problem formulation. The linear input and output constraints can be write as:

$$\begin{aligned} \mathbf{u}_{\text{min}} \leq \mathbf{u}(k) \leq \mathbf{u}_{\text{max}}, \\ \mathbf{y}_{\text{min}} \leq \mathbf{y}(k) \leq \mathbf{y}_{\text{max}}. \end{aligned} \tag{B.22}$$

States are constrained as $\mathbf{x}(k) \in \mathbb{X}$, $k = 1, 2, \ldots N_o$, where $\mathbb{X}$ is a polyhedral set of the form

$$\mathbb{X} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{L}_k \mathbf{x}(0) \leq \mathbf{W}_k\}, \quad k = 1, 2, \ldots N_o, \tag{B.23}$$

where $\mathbf{L}_k$, $\mathbf{W}_k$ explained below. The constraints (B.22)–(B.23) can be written as linear constraints on $\mathbf{u}$ of the compact form

$$\mathbf{L}\mathbf{U} \leq \mathbf{W} \tag{B.24}$$

where

$$\mathbf{L} = \begin{bmatrix} \mathbf{I}_{N_o m} \\ \boldsymbol{\Psi} \\ -\mathbf{I}_{N_o m} \\ -\boldsymbol{\Psi} \\ \tilde{\mathbf{L}} \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{U}_{\max} \\ \mathbf{Y}_{\max} \\ \mathbf{U}_{\min} \\ \mathbf{Y}_{\min} \\ \tilde{\mathbf{W}} \end{bmatrix}. \tag{B.25}$$

In (B.25) $\mathbf{I}_{N_o m}$ is the $N_o m \times N_o m$ identity matrix (where $N_o = N_p$ is the control horizon and m is the number of inputs). $\boldsymbol{\Psi}$ is the $(N_o - 1)m \times N_o m$ size matrix:

$$\boldsymbol{\Psi} = \begin{bmatrix} \mathbf{CB} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{CAB} & \mathbf{CB} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \\ \mathbf{CA}^{N_o-1}\mathbf{B} & \mathbf{CA}^{N_o-2}\mathbf{B} & \ldots & \mathbf{CB} \end{bmatrix}, \tag{B.26}$$

$$\tilde{\mathbf{L}} = \text{diag}\{\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_{N_o}\}\boldsymbol{\Gamma}, \tag{B.27}$$

where $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_{N_o}$ are the state constraint matrices given in (B.23) while $\boldsymbol{\Gamma}$ is given in (B.18). The vectors develop $\mathbf{W}$ are defined as below:

$$\mathbf{U}_{\max} = \begin{bmatrix} \mathbf{u}_{\max} \\ \mathbf{u}_{\max} \\ \vdots \\ \mathbf{u}_{\max} \end{bmatrix}, \quad \mathbf{U}_{\min} = \begin{bmatrix} \mathbf{u}_{\min} \\ \mathbf{u}_{\min} \\ \vdots \\ \mathbf{u}_{\min} \end{bmatrix},$$

$$\mathbf{Y}_{\max} = \begin{bmatrix} \mathbf{y}_{\max} - \mathbf{C}\mathbf{A}\mathbf{x}(0) - \mathbf{d}_1 \\ \vdots \\ \mathbf{y}_{\max} - \mathbf{C}\mathbf{A}^{N_p}\mathbf{x}(0) - \mathbf{d}_{N_p} \end{bmatrix}, \quad \mathbf{Y}_{\min} = \begin{bmatrix} -\mathbf{y}_{\min} + \mathbf{C}\mathbf{A}\mathbf{x}(0) + \mathbf{d}_1 \\ \vdots \\ -\mathbf{y}_{\min} + \mathbf{C}\mathbf{A}^{N_p}\mathbf{x}(0) + \mathbf{d}_{N_p} \end{bmatrix},$$

$$\tilde{\mathbf{W}} = -\text{diag}\{\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_{N_o}\} \begin{bmatrix} \mathbf{A}\mathbf{x}(0) \\ \vdots \\ \mathbf{A}^{N_p}\mathbf{x}(0) \end{bmatrix} + \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_{N_p} \end{bmatrix},$$

where $\mathbf{x}(0)$ is the initial state, $\mathbf{u}_{\max}, \mathbf{u}_{\min}, \mathbf{y}_{\max}, \mathbf{y}_{\min}$ are the vectors of constraint defined in (B.22) and $\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_{N_o}, \mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{N_o}$ are the matrices and vectors of the state constraint polyhedra (B.23).

Using the above formalism, one can express the problem of minimising (B.20) subject to the equality and inequality constraints in the form of QP problem as follows:

$$\begin{aligned} &\min_{\mathbf{U}} \frac{1}{2}\mathbf{U}^{\mathsf{T}}\mathbf{H}\mathbf{U} + \mathbf{U}^{\mathsf{T}}[\mathbf{F}\mathbf{x}(0) + \mathbf{G}\Delta\mathbf{D}] \\ &\text{subject to:} \\ &\mathbf{L}\mathbf{U} \leq \mathbf{W}. \end{aligned} \tag{B.28}$$

The optimal solution $\mathbf{u}^*(x)$ to (B.28) is then:

$$\mathbf{u}^*(x_0) = \arg \min_{\mathbf{L}\mathbf{U} \leq \mathbf{W}} \frac{1}{2}\mathbf{U}^{\mathsf{T}}\mathbf{H}\mathbf{U} + \mathbf{U}^{\mathsf{T}}[\mathbf{F}\mathbf{x}(0) + \mathbf{G}\Delta\mathbf{D}]. \tag{B.29}$$

Once the open-loop QP problem (B.29) is solved from the known initial $\mathbf{x}(0)$ and predicted disturbances $\mathbf{d}(k), k = 0, 1, \ldots, N_p - 1$, the rolling horizon scheme applies, at the current time k, only the first control move, formed by the first m components of the optimal vector $\mathbf{U}^*(\mathbf{x}_0)$ in (B.29). This yields a control law of the form

$$\mathbf{u}(k) = \mathcal{M}[\mathbf{x}(k), \mathbf{d}(\kappa)], \quad \kappa = k, k+1, \ldots, k+N_p-1 \tag{B.30}$$

where $\mathbf{x}(k) = \mathbf{x}_0, k = 0, \ldots, N_o - 1$ is the current state of the system and $\mathcal{M}$ is a linear mapping from the state and disturbance spaces to control. Then the whole procedure is repeated at the next time instant, with the optimisation horizon kept constant. Note that the analytical solution (B.21) for the unconstrained problem is of particular interest; given that the optimal solution for

the constrained problem has a similar form in a region of the state space where the state of the system vanishes, i.e. $\Delta\mathbf{x} = \mathbf{0}$ or $\mathbf{x} = \hat{\mathbf{x}}$.

Bibliography

[1] K. Aboudolas and N. Geroliminis. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. *Transportation Research Part B*, 55:265–281, 2013.

[2] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos. Store-and-forward based methods for the signal control problem in large-scale congested urban road networks. *Transportation Research Part C*, 17(2):163–174, 2009.

[3] K. Aboudolas, M. Papageorgiou, A. Kouvelas, and E. Kosmatopoulos. A rolling-horizon quadratic-programming approach to the signal control problem in large-scale congested urban road networks. *Transportation Research Part C*, 18(5):680–694, 2010.

[4] A. Allström, J. Barceló, J. Ekström, E. Grumert, D. Gundlegård, and C. Rydergren. Traffic management for smart cities. In V. Angelakis, E. Tragos, H. C. Pöhls, A. Kapovits, and A. Bassi, editors, *Designing, Developing, and Facilitating Smart Cities: Urban Design to IoT Solutions*, pages 211–240. Springer International Publishing, 2017.

[5] E. Alpaydin. *Introduction to machine learning*. MIT press, 2009.

[6] L. Ambühl and M. Menendez. Data fusion algorithm for macroscopic fundamental diagram estimation. *Transportation Research Part C*, 71:184–197, 2016.

[7] K. Ampountolas and A. Kouvelas. Real-time estimation of critical vehicle accumulation for maximum network throughput. In *2015 American Control Conference*, pages 2057–2062, 2015.

[8] K. Ampountolas, Nan Zheng, and N. Geroliminis. Robust control of bi-modal multi-region urban networks: An LMI optimisation approach. In *17th International IEEE Conference on Intelligent Transportation Systems*, pages 489–494, Oct 2014.

[9] K. Ampountolas, N. Zheng, and N. Geroliminis. Macroscopic modelling and robust control of bi-modal multi-region urban road networks. *Transportation Research Part B*, 104:616–637, 2017.

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

[11] W. M. Bolstad and J. M. Curran. *Introduction to Bayesian Statistics*. Wiley-Blackwell, 3rd edition, 2016.

[12] F. Borrelli, A. Bemporad, and M. Morari. *Predictive control for linear and hybrid systems*. Cambridge University Press, Cambridge, UK, 2017.

[13] D. Bretherton, G. Bowen, and K. Wood. Effective urban traffic management and control: Recent developments in SCOOT. In *82nd TRB Annual Meeting*, Washington, DC, USA, 2003.

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

[15] M. Cannon. C21 model predictive control, 2015. Lecture Notes: http://www.eng.ox.ac.uk/~conmrc/mpc/notes.html.

[16] N. Chiabaut. Evaluation of a multimodal urban arterial: The passenger macroscopic fundamental diagram. *Transportation Research Part B*, 81(2):410–420, 2015.

[17] R. Church and C. ReVelle. The maximal covering location problem. *Papers of the Regional Science Association*, 32(1):101–118, 1974.

[18] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. The MIT Press, Cambridge, MA, 3rd edition, 2009.

[19] T. Courbon and L. Leclercq. Cross-comparison of macroscopic fundamental diagram estimation methods. *Procedia - Social and Behavioral Sciences*, 20:417–426, 2011.

[20] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley, New Jersey, 2nd edition, 2006.

[21] A. Csikós, T. Tettamanti, and I. Varga. Nonlinear gating control for urban road traffic network using the network fundamental diagram. *Journal of Advanced Transportation*, 49 (5):597–615, 2015.

[22] J. Current, M. Daskin, and D. Schilling.

[23] C. F. Daganzo. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B*, 41(1):49–62, 2007.

[24] C. F. Daganzo and N. Geroliminis. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B*, 42(9):771–781, 2008.

[25] M. S. Daskin. *Network and discrete location: Models, algorithms and applications*. John Wiley and Sons, 1995.

[26] D. de Jong, V. L. Knoop, and S. P. Hoogendoorn. The effect of signal settings on the macroscopic fundamental diagram and its applicability in traffic signal driven perimeter control strategies. In *16th International IEEE Conference on Intelligent Transportation Systems*, pages 1010–1015, Oct 2013.

[27] C. Diakaki, M. Papageorgiou, and K. Aboudolas. A multivariable regulator approach to traffic-responsive network-wide signal control. *Control Engineering Practice*, 10(2):183–195, 2002.

[28] J. Du, H. Rakha, and V. V. Gayah. Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. *Transportation Research Part C*, 66:136–149, 2016.

[29] L. C. Edie. Discussion of traffic stream measurements and definitions. In J. Almond, editor, *2nd International Symposium on the Theory of Traffic Flow*, pages 139–154, OECD, Paris, France, 1963.

[30] R. Z. Farahani, N. Asgari, N. Heidari, M. H., and M. Goh. Covering problems in facility location: A review. *Computers & Industrial Engineering*, 62(1):368–407, 2012.

[31] K. Fox, P. Franklin, and F. Montgomery. PRIMAVERA final report. http://www.its.leeds.ac.uk/projects/primavera/finrep.html.

[32] V. V. Gayah, X. Gao, and A. S. Nagle. On the impacts of locally adaptive signal control on urban network stability and the macroscopic fundamental diagram. *Transportation Research Part B*, 70:255–268, 2014.

[33] N. Geroliminis and B. Boyaci. The effect of variability of urban systems characteristics in the network capacity. *Transportation Research Part B*, 46(10):1607–1623, 2012.

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

[35] N. Geroliminis and D. M. Levinson. Cordon pricing consistent with the physics of overcrowding. In W. H. K. Lam, S. C. Wong, and H. K. Lo, editors, *Transportation and Traffic Theory 2009: Golden Jubilee: Papers selected for presentation at ISTTT18, a peer reviewed series since 1959*, pages 219–240. Springer US, Boston, MA, 2009.

[36] N. Geroliminis and J. Sun. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B*, 45(3):605–617, 2011.

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

[38] N. Geroliminis, N. Zheng, and K. Ampountolas. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. *Transportation Research Part C*, 42: 168–181, 2014.

[39] J. W. Godfrey. The mechanism of a road network. *Traffic Engineering& Control*, 11(7): 323–327, 1969.

[40] G. Goodwin, M. M. Seron, and J. A. de Don. *Constrained control and estimation: An optimisation approach*. Springer-Verlag, London, UK, 2005.

[41] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray. On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, 42(2): 251–260, 2006.

[42] J. Haddad. Robust constrained control of uncertain macroscopic fundamental diagram networks. *Transportation Research Part C*, 59:323–339, 2015.

[43] J. Haddad. Optimal coupled and decoupled perimeter control in one-region cities. *Control Engineering Practice*, 61:134–148, 2017.

[44] J. Haddad. Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics. *Transportation Research Part B*, 96:1–25, 2017.

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

[46] J. Haddad and A. Shraiber. Robust perimeter control design for an urban region. *Transportation Research Part B*, 68:315–332, 2014.

[47] J. Haddad, N. Geroliminis, and M. Ramezani. Cooperative traffic control of a mixed network with two urban regions and a freeway. *Transportation Research Part B*, 54:17–36, 2013.

[48] M. Hajiahmadi, V. L. Knoop, B. De Schutter, and H. Hellendoorn. Optimal dynamic route guidance: A model predictive approach using the macroscopic fundamental diagram. In *16th International IEEE Conference on Intelligent Transportation Systems*, pages 1022–1028, 2013.

[49] M. Hajiahmadi, J. Haddad, B. De Schutter, and N. Geroliminis. Optimal hybrid perimeter and switching plans control for urban traffic networks. *IEEE Transactions on Control Systems Technology*, 23(2):464–478, 2015.

[50] S. L. Hakimi. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research*, 13(3):462–475, 1965.

[51] R. Herman and I. Prigogine. A two-fluid approach to town traffic. *Science*, 204:148–151, 1979.

[52] G. E. Hovland and B. J. McCarragher. Dynamic sensor selection for robotic systems. In *Proceedings of International Conference on Robotics and Automation*, volume 1, pages 272–277, 1997.

[53] P B. Hunt, D. I. Robertson, R. D. Bretherton, and M. C. Royle. The SCOOT on-line traffic signal optimisation technique. *Traffic Engineering and Control*, 23(4):190–192, 1982.

[54] Y. Ji and N. Geroliminis. On the spatial partitioning of urban transportation networks. *Transportation Research Part B*, 46(10):1639–1656, 2012.

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

[56] M. Keyvan-Ekbatani, M. Papageorgiou, and I. Papamichail. Urban congestion gating control based on reduced operational network fundamental diagrams. *Transportation Research Part C*, 33:74–87, 2013.

[57] M. Keyvan-Ekbatani, R. C. Carlson, V. L. Knoop, S. P. Hoogendoorn, and M. Papageorgiou. Queuing under perimeter control: Analysis and control strategy. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems*, pages 1502–1507, 2016.

[58] V. Knoop and S. Hoogendoorn. Empirics of a generalized macroscopic fundamental diagram for urban freeways. *Transportation Research Record*, 2391:133–141, 2013.

[59] V. L. Knoop, S. Hoogendoorn, and J. W. C. Van Lint. Routing strategies based on the macroscopic fundamental diagram. *Transportation Research Record*, 2315(1):1–10, 2012.

[60] A. Kouvelas, K. Aboudolas, M. Papageorgiou, and E.B. Kosmatopoulos. A hybrid strategy for real-time traffic signal control of urban road networks. *IEEE Transactions on Intelligent Transportation Systems*, 12(3):884–894, 2011.

[61] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008.

[62] B. Kulcsár, K. Ampountolas, and A. Dabiri. Single-region robust perimeter traffic flow control. In *2015 European Control Conference*, pages 2628–2633, 2015.

[63] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[64] G. Laporte, S. Nickel, and F. Saldanha da Gama, editors. *Location Science*. Springer, Cham, Switzerland, 2015.

[65] L. Leclercq, N. Chiabaut, and B. Trinquier. Macroscopic fundamental diagrams: A cross-comparison of estimation methods. *Transportation Research Part B*, 62:1–12, 2014.

[66] A. Leonardo, L. Arnold, and P. Roland. Information theory-based approach for location of monitoring water level gauges in polders. *Water Resources Research*, 46(3), 2010.

[67] P. R. Lowrie. SCATS: the Sydney co-ordinated adaptive traffic system–Principles, methodology, algorithms. In *IEEE International Conference on Road Traffic Signalling*, London, England, 1982.

[68] J. M. Maciejowski. *Predictive control with constraints*. Prentice Hall, Harlow, UK, 2002.

[69] H. Mahmassani, J.C. Williams, and R. Herman. Performance of urban traffic networks. In N.H. Gartner and N.H.M. Wilson, editors, *10th International Symposium on Transportation and Traffic Theory*, Elsevier, Amsterdam, The Netherlands, 1987.

[70] A. D. May, R. Liu, S. P. Shepherd, and A. Sumalee. The impact of cordon design on the performance of road pricing schemes. *Transport Policy*, 9(3):209–220, 2002.

[71] A. Mazloumian, N. Geroliminis, and D. Helbing. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society A*, 368(1928):4627–4647, 2010.

[72] A. S. Nagle and V. V. Gayah. Accuracy of networkwide traffic states estimated from mobile probe data. *Transportation Research Record*, 2421(1):1–11, 2014.

[73] J. Ortigosa, M. Menendez, and H. Tapia. Study on the number and location of measurement points for an mfd perimeter control scheme: a case study of zurich. *EURO Journal on Transportation and Logistics*, 3(3):245–266, 2014.

[74] C. Papadimitriou. Optimal sensor placement methodology for parametric identification of structural systems. *Journal of Sound and Vibration*, 278(4):923–947, 2004.

[75] M. Ramezani, J. Haddad, and N. Geroliminis. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transportation Research Part B*, 74:1–19, 2015.

[76] M. Saeedmanesh and N. Geroliminis. Clustering of heterogeneous networks with directional flows based on "Snake" similarities. *Transportation Research Part B*, 91:250–269, 2016.

[77] M. Saeedmanesh and N. Geroliminis. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. *Transportation Research Part B*, 105: 193–211, 2017.

[78] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, (27):379–423, 1948.

[79] L.V Snyder. Covering problems. In H.A Eiselt and V. Marianov, editors, *Foundations of location analysis*. Springer Science & Business Media, 2011.

[80] A. Sumalee, S. P. Shepherd, and A. May. Road user charging design: dealing with multi-objectives and constraints. *Transportation*, 36(2):167–188, 2009.

[81] M. S. Terman, N. M. Kojouri, and H. Khalafi. Optimal placement of fixed in-core detectors for Tehran research reactor using information theory. *Progress in Nuclear Energy*, 106: 300–315, 2018.

[82] C. Toregas, R. Swain, C. ReVelle, and L. Bergman. The location of emergency service facilities. *Operations Research*, 19(6):1363–1373, 1971.

[83] T. Tsubota, A. Bhaskar, and E. Chung. Macroscopic fundamental diagram for Brisbane, Australia. *Transportation Research Record*, 2421(1):12–21, 2014.

[84] R. A. Vincent and R. E. Layfield. Nottingham zones and collar study–Overall assessment. In *Transport and Road Research Laboratory Report 805*, Crowthorne, 1977.

[85] F. Viti, M. Rinaldi, F. Corman, and C. M. J. Tampère. Assessing partial observability in network sensor location problems. *Transportation Research Part B*, 70:65–89, 2014.

[86] K. Wood, D. Bretherton, A. Maxwell, K. Smith, and G. Bowen. Improved traffic management and bus priority with SCOOT (TRL staff paper PA 3860/02). In *Transport Research Laboratory*, London, UK, 2002.

[87] M. Yildirimoglu, M. Ramezani, and N. Geroliminis. Equilibrium analysis and route guidance in large-scale networks with MFD dynamics. *Transportation Research Part C*, 59: 404–420, 2015.

[88] L. Zhang, T. M. Garoni, and J. de Gier. A comparative study of Macroscopic Fundamental Diagrams of arterial road networks governed by adaptive traffic signal systems. *Transportation Research Part B*, 49:1–23, 2013.

[89] F. Zhao and L. Guibas. *Wireless Sensor Networks: An Information Processing Approach*. Morgan Kaufmann, San Mateo, CA, 2004.

[90] N. Zheng and N. Geroliminis. Modeling and optimization of multimodal urban networks with limited parking and dynamic pricing. *Transportation Research Part B*, 83:36–58, 2016.

[91] N. Zheng, R. A. Waraich, K. W. Axhausen, and N. Geroliminis. A dynamic cordon pricing scheme combining the Macroscopic Fundamental Diagram and an agent-based traffic model. *Transportation Research Part A*, 46(8):1291–1303, 2012.

[92] A. Zockaie, M. Saberi, and R. Saedi. A resource allocation problem to estimate network fundamental diagram in heterogeneous networks: Optimal locating of fixed measurement points and sampling of probe trajectories. *Transportation Research Part C*, 86:245–262, 2018.