The Dynamics and the Uncertainty of Delays at Signals

Francesco Viti

Delft University of Technology, 2006

Cover illustration: *HippoDG*

The Dynamics and the Uncertainty of Delays at Signals

Proefschrift

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft

op gezag van de Rector Magnificus prof. dr. ir. J.T. Fokkema,

voorzitter van het College van Promoties,

in het openbaar te verdedigen op maandag 13 november 2006 om 10.00 uur

door Francesco VITI

civiel ingenieur

geboren te Matera (Italië)

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. H.J. van Zuylen

Samenstelling Promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. H.J. van Zuylen	Technische Universiteit Delft, promotor
Prof. dr. ir. P.H.L. Bovy	Technische Universiteit Delft
Prof. dr. ir. S.P. Hoogendoorn	Technische Universiteit Delft
Prof. dr. H.C. Tijms	Vrije Universiteit Amsterdam
Prof. ir. L.H. Immers	Katholieke Universiteit Leuven, België
Prof. Dr-Ing. W. Brilon	Ruhr-Universität Bochum
Dr. P. Olszewski	Nanyang TU, Singapore

This thesis is the result of a Ph.D. study carried out from November 2001 till November 2006 at the Delft University of Technology.

TRAIL Thesis Series nr. T2006/7, The Netherlands TRAIL Research School

Published and distributed by: TRAIL Research School

TRAIL
P.O. Box 5017
2600 GA Delft, The Netherlands
Phone: +31 (0) 15 278 6046
Fax: +31 (0) 15 278 4333
E-mail: info@rsTRAIL.nl

ISBN 90-5584-078-5

Printed in The Netherlands

“If you are in a rush the traffic lights
are always red when you get to them”.

- Murphy’s Law of Traffic Lights

Preface

I'll go immediately to the point, I swear. During these 5 years of Ph.D. I was constantly advised to keep focus on my own research, and to avoid useless text in my papers. Now, after more than one year of writing and deleting, I am repenting not to have learnt the lesson earlier. This thesis could have been two times thicker and dealing with traffic flow operation issues, behavioral studies, game theory, etc. Thanks to the constant remarks of a few colleagues, and especially of my supervisor, Professor Henk van Zuylen, I was able to produce this book, which deals with only one specific problem: the queuing process at signalized intersections.

This thesis describes in fact the progress made in the modeling of queues and delays at traffic signals and discusses the limitations of these models in describing the stochastic and dynamic behavior of these service systems. Starting from a well-established theory in operations research, the renewal theory of Markov Chains, which has been applied in the past to investigate and analyze the dynamic behavior of overflow queues at fixed time signals, we developed and integrated within this modeling framework a probabilistic formulation also for the queue behavior within each cycle. This model enables one to deal with queues using a continuous time approach, and it describes the effect of the variability of the arrivals in the service time process, which reflects into the variability of the delay caused by the signal operation. The flexibility of this modeling framework allows its application in more sophisticated service systems, i.e. paired intersections, multiple service points and demand-responsive signals.

During the research, several people have contributed to its successful ending. Firstly, I would like to thank AVV to financially support this Ph.D. project and my supervisor, Professor Henk van Zuylen to believe in my potentials, even if sometimes our lines of thought were not completely matching. Thanks also to Dr. Yu Sen Chen for giving a critical view on a few articles and for allowing the use of AIMSUN at DHV. Many thanks go also to the TRAIL Research School, which to my opinion makes a good job in guaranteeing an intense knowledge exchange between Ph.D. students belonging to this school and in promoting their research outside.

I am particularly grateful also to two exemplary professors, Professor Piet Bovy and Professor Serge Hoogendoorn, who well describe the continuity of research beyond the walls of our department. Their comments and remarks during these years and their support during the writing period have undoubtedly increased the quality and solidity of the issues described and relaxed in this book. The success of our research group starts from very good mentors, but flourishes with the enthusiasm and the ensemble of the researchers that formed it during these years. We are now too many to acknowledge them all and thank them one by one, but every one of them knows how much I appreciated to spend my days with them talking about work, football, movies etc. and to play ping pong sharing always good laughs and a lot of emotions. One special hug I want to reserve to Dr. Hans van Lint, who convinced me to move to the Netherlands. Without you I could never appreciate this country so much!

Thanks also to my family in the Netherlands, my *HippoDG* housemates, and my friends, the *buitenlanders*. Finally I would like to thank my parents and my brother, who I missed so much during these 5 years and whom I am going to miss a lot every day I won't spend with them in the coming years.

And now, Mariangela, I would like to start a new project together with you…

Contents

Notation

Acronyms

ITS	Intelligent Transportation Systems
DTM	Dynamic Traffic Management
DTA	Dynamic Traffic Assignment
DNL	Dynamic Network Loading
ADAS	Advanced Driver Assistance Systems
ATIS	Advanced Travel Information Systems
RUT	Random Utility Theory
HCM	Highway Capacity Manual
PCE	Passenger car equivalent
LoS	Level of Service
TSL	Travel Simulator Laboratory
CoV	Coefficient of variation
FIFO	First In First Out
LIFO	Last In Last Out

Mathematical/statistical operators

$E[.]$	Expectation value
$\sigma[.]$	Standard deviation
$\sigma^2[.]$	Variance
$\Pr(.)$	Probability
$[.]$	Integer value of a real number

Travel time attributes of the TSL experiment

n	Repetition number
i	Index for a respondent
k	Index for a route
$\mathrm{I}_{ik}(n)$	Score index in the TSL
$t_{ik}^{early}(n)$	Time (in minutes) the respondent has arrived (eventually) early at destination
$t_{ik}^{driving}(n)$	Driving time (in minutes)

$t_{ik}^{late}(n)$	Time (in minutes) the respondent has arrived (eventually) late at destination
$\alpha_{ik}, \beta_{ik}, \gamma_{ik}$	Value of early arrival, driving time and late arrival
$U_{ik}(n)$	Utility value
$\mathbf{X}_{ik}(n)$	Vector of decisional variables
ASC_{ik}	Alternative specific constant
$\boldsymbol{\beta}$	Vector of relative values for the decisional variables
ε	Error term in the utility function

Traffic control parameters

t_g	Effective green time (in seconds)
t_r	Effective red time
t_C	Cycle time

Traffic flow and performance variables

t	Time
q	Arrival rate (veh/h)
s	Saturation flow rate (veh/h)
W	Delay (s)
W_1	Uniform stopped delay per vehicle (s/veh)
W_2	Incremental, or random stopped delay (s/veh)
W_3	Initial queue delay
PF	Progression factor, to account for signal coordination in the HCM
I_f	Adjustment factor to account for the filtering effect in the HCM
k	Adjustment factor for vehicle actuated controls
u,t	Parameters for the initial queue delay in the HCM
k_f	Progression factor, to account for signal coordination in the Canadian Guide
m	Adjustment factor to account for the filtering effect in the Australian Capacity Guide
x_0	Minimum value of the degree of saturation to compute the overflow queue
Q_O	Overflow queue length (in pcu)
Q_t	Overflow queue length (in pcu) at time t
$Q(0)$	Initial queue length
$A(t)$	Cumulative number of arrivals during the period $[0,t]$
$D(t)$	Cumulative number of departures during the period $[0,t]$
$Q(t)$	Queue length at time t

I	Index of dispersion
Z_2	Total delay experienced during green when the cycle t_C is infinite
$c = s \cdot t_g / t_C$	Signal capacity (veh/h)
$x = q / c$	Degree of saturation
$y = q / s$	Flow-to-saturation flow ration
Δc	Reserve capacity in one cycle
$H(\mu)$	Adjustment factor for the overflow queue where $\mu = \frac{s \cdot t_g - q \cdot t_C}{\sqrt{I \cdot s \cdot t_g}}$
P_{coord}	Probability of a vehicle to arrive during green in a coordinated signal
f_p	Progression adjustment factor for signal coordination
I_a	Dispersion index of arrivals in signal coordination in the Van As model
F	Filtering factor in the Van As model
B	Dispersion index of departures in the Van As model
L	Total lost time in a cycle
x_c	Critical volume to capacity ratio in the HCM
K_1, K_2	Discount factors in the modified Webster function

Markov chain notation

Q	Queue length
$Q_{\max}$	Maximum queue length (veh)
$q_{\max}$	Maximum number of arrival at the intersection approach within a cycle
$d_{\max}$	Maximum number of departures within a cycle
Q_{ij}	Transition matrix
$\Pr(Q_o = j, t)$	Probability of overflow queue being *j* at time *t*
D_1, D_2, D_3	Three delay components of the Olszewski's delay model
$q_{\Delta t}$	Number of arrivals during time Δt

Van Zuylen-Viti model

Q_{lin}	Linear deterministic function
$Q_{\exp}$	Exponential function
Q_{MC}	Queue simulated with the Markov Chain process
α, β, γ	Calibration parameters for the Van Zuylen-Viti model
μ, To	Calibration parameters for the parameter α

Lane changing variables

$\Psi_{a \to b}$	Probability of intention to change lane
$\eta_{a \to b}$	Effective lane changes
$Q_{spillback}$	Number of vehicles which can be contained in a flare

Vehicle actuated control variables

$\bar{\tau}$	Green time unit
$g^r(\tau)$	Green time portion to serve vehicles arriving during the red phase
$g_{\min}$, $g_{\max}$	Minimum and maximum green time values
Q^r	Number of vehicles queuing during the red phase
r	Red time of approach *i*
Q_i^g	Vehicles queuing at the back of the queue during the green phase
g_i^g	Green time portion to serve the vehicles arriving while green
g_i^Q	Total green time for a full queue discharge
Q_i	Total queue
g_i^e	Green time extension
g_i^{tot}	Total green time given to approach *i*
Q_O^i	Overflow queue length
TL	Total time lost in one signal cycle

1

Introduction

Traffic congestion on freeways and urban areas causes nowadays enormous economic losses worldwide. In the Netherlands the Dutch Ministry of Transport (AVV) has estimated, for the year 1997, a loss due to traffic congestion of 1.7 billion Gulden (nearly 1 billion €) (AVV 1998). Every day the traffic on the Dutch motorways produces serious delays to the drivers due to congestion, represented in an average working day by 200km of queues (Bovy 2001). To give another example, a large-scale report for the American highways (the 2005 Urban Mobility Report), involving 81 major cities in the U.S., estimated for the year 2003 an average yearly loss for a commuter of 47 hours of delays (against only 16 hours in 1982) and 28 gallons of extra fuel consumed, resulting in an average loss of $722 per commuter per year. The worst congestion levels increased from 12% to 40% in the peak period travel in the largest cities and uncongested periods decreased from 70% of the day to only 33% in the period 1982-2003 (Schrank 2005).

Congestion levels are therefore becoming more and more severe and peaks of the demand are involving longer time periods. Recent policies, meant to relieve this traffic, motivate the development of new transport management strategies with the objective of a more efficient utilization of the system. A large contribution in this direction is given by the partnership of computer technology and scientific research through the development of technologically advanced systems (usually referred to as *Intelligent Transportation Systems*, or simply *ITS*) to support the driver and manage the traffic. The objective of these systems is to guarantee safety and comfortable conditions to the drivers, and, whether it is possible, to reduce congestion. Reduction of congestion is therefore achieved by means of a more efficient use of the existing physical capacity, i.e. by strategies that directly affect the demand for travel (e.g. information, route guidance systems, telecommuting, pricing) or that modify the capacity dynamically according to the actual need for road space (e.g. signal control, speed control, ramp meters, incident management). It is expected that the implementation of these systems will provide the following effects (Van Zuylen 2003):

- Motorways: 20 - 30% less queues
- Rural roads: 2 - 20%
- Urban roads: 10 - 20% less delays.

These represent only approximate estimates, since they do not consider either future growth of the demand and its response to changes in the supply system (e.g. induced demand, re-routing, shift of congestion to other parts of the network etc.). The assessment is done by simulation programs, where the traffic flow propagation on the network is simulated as realistic as possible and the demand is estimated using iterative procedures like the traffic assignment. Therefore, analytic models, which require small computation times, are very appealing for traffic planners. On the other hand, traffic flow models should give estimates that are consistent with real life, i.e. they should deal with the dynamic and probabilistic character of traffic processes.

This issue affects especially urban networks, where the drivers have usually more routing possibilities. Researchers pointed out that disagreement between traffic flow models and actual travel times and choices in urban networks are for a consistent part due to poor estimation of delays at intersections, especially during congested conditions (e.g. see (Rouphail 2000)). In these conditions the dynamic behavior of delays is strongly dependent on the queuing process caused by the control mechanism. There is still lack of a complete understanding of the dynamic behavior of queues, and large uncertainty still characterizes its predictability.

This thesis provides a methodology to analyze the dynamic and probabilistic character of the queuing process at signalized intersections and its effect on the drivers' delay; this methodology considers explicitly the variability of the state variables (i.e. demand, capacity etc.). Particular interest is given in this thesis to the delay estimation problem at signalized intersections because of its fundamental role in the total delay drivers experience in urban areas and in the estimation of the network performance, e.g. in the estimation of the level of service.

The probabilistic approach described in this thesis has been adopted for various purposes: 1) to validate and compare previously published formulas based on static (or quasi-static) assumptions, 2) to give insight into the dynamic and stochastic character of queues and delays depending on the random nature of the transportation system, 3) to inspire new formulations, which overcome the limitations of the formulas currently applied in planning and design of signalized network problems and 4) to model the queuing process at signalized intersections in complex systems like with multiple lanes and dynamic arrival-dependent controls. The probabilistic modeling of delay processes, i.e. the *mesoscopic approach,* allows the estimation of performance measures adopted in most planning and design problems, e.g. level of service or travel time, and the estimation of their variability in time.

This introductory chapter is organized as follows. Section 1.1 describes the problem and briefly defines the area of research covered by this thesis. In Section 1.2 the objectives and scope of the research are described. The implications and the contributions of this thesis to the state-of-the-art in traffic flow modeling at traffic controls are summarized in Section 1.3 while Section 1.4 gives an outline of this thesis.

1.1 Problem formulation

Traffic congestion on a road section is in general caused by part of the demand that exceeds the available capacity during a certain time period. A way to reduce this problem is intervening on the infrastructure, e.g. by adding new lanes or building new roads. This type of intervention implies however high costs and long periods of inconvenience for the traffic due to the necessary road works. Moreover, this intervention may be beneficial for a small period of the day, e.g. during peak periods, while it may represent a waste during off-peak periods. As an alternative, transport managers can improve the network conditions by using the available network infrastructure more efficiently. Some of these alternative management strategies are referred to as *Dynamic Traffic Management* (*DTM*) measures. These strategies are designed with the objective of using the network infrastructure efficiently, while keeping high levels of safety and comfort. This objective is obtained by adapting (dynamically) the road capacity to the demand (e.g. dynamic traffic control, speed control, ramp metering etc.) or vice versa (e.g. by using pricing policies or by guiding the drivers towards alternative routes). The design and planning of such systems require models that predict the expected benefits on the traffic system. A good estimation or prediction of network flows together with the corresponding costs (e.g. delay, fuel consumption, air pollution etc.) as a function of the applied DTM strategies is very important for an optimal set-up of such measures according to the policy objectives, and therefore for acceptance by the road authorities and the road users.

Transport planning and design have been historically concerned with travel behavior and the transport system in some nominally "typical" conditions (Clark 2005). Therefore, these problems have been typically solved using analytic travel time models, based on some average conditions of traffic. The reasons for this approach are various: they require small computation times, therefore they are suited for optimization algorithms or for iterative procedures, and they directly relate the performance measures to the state and control variables, allowing e.g. sensitivity analyses. An alternative approach to the analytic approach is using simulation; while microscopic simulation models simulate the traffic in a very detailed way, i.e. by simulating the movement and the characteristics of each vehicle, macroscopic simulation models need less computational effort, since they simulate the traffic flow process at a higher aggregation level.

The estimation and prediction of travel times is, however, largely affected by the complex structure of urban networks and by the dynamic and stochastic behavior of demand and supply systems. From this perspective both analytic and simulation based approaches have limitations. Analytic models and macroscopic simulation models lack in catching these dynamic and stochastic effects because of their relatively simple model structure, while microscopic models simulate only one of the possible situations that can occur and several simulations are needed to obtain long-term estimates.

This problem affects in particular the modeling of delays at signalized intersections. An intersection operates at different traffic conditions and it may operate at level of services accepted by the policy makers only for a fraction of the day. This does not occur when it should serve the upmost part of the total daily demand, i.e. at peak periods. The scientific forum agrees that large improvements are still needed in the modeling of delays, above all, because of the behavior of queues forming and dissipating within a cycle and cycle-by-cycle at controlled intersections and their effects to the capacity and the throughput of a network, such as spill back. In particular, control delay models are lacking of a good queuing formula, which enables one to fully catch the dynamic and the stochastic behavior of traffic. Contribution to this gap is necessary due to a lack of a queuing formula that gives correct estimates of the dynamic effects of congestion.

To overcome this problem most of these approximate formulas have been modified based on heuristics. For example, Webster's formula (see Chapter 3, Formula (3.9)) was corrected to fit better simulation data using a heuristic correction term, which does not have any theoretical meaning. Different heuristics were used first by Kimber and Hollis (Kimber 1979) and later by Akcelik in both his time-dependent travel time functions ((Akcelik 1980), (Akcelik 1991)), who applied the coordinate transformation technique to derive time dependent formulations for the expectation value of queues and delays from their exact expressions in the static context.

The heuristic foundation of these models implies that there is still no clear insight into the real dynamic behavior of these measures. If there is no clear insight into the delay and queuing behavior through a theoretically sound methodology, then all models will be deficient in catching the real dynamics of traffic. Modeling queues and delays through probabilistic models can help in better understanding this behavior and inspire new approximate analytic formulas, as it is demonstrated in this thesis (Chapter 6). Analyzing traffic at signals using Markov Chains is not a completely new approach; a few studies can in fact be enumerated (e.g. (Van Zuylen 1985), (Olszewski 1990) among others, see

Chapters 3 and 4 for a more detailed descriptions of these studies) and very few used this methodology to derive approximate analytic expressions (e.g. (Brilon 1990), (Wu 1990), (Fu 2000)).

Probabilistic models are an alternative to analytic and microscopic simulation models. These models use true macroscopic relationships between state, control variables and the resulting performance measures, and they assume these variables as statistically distributed according to a known probability distribution function. Consequently, also the performance measures are calculated in a probabilistic fashion. This class of models is increasingly gaining the attention of the traffic analysts, since it can catch the stochastic character of the performance measures, i.e. their variability. Among this class, probabilistic models based on renewal processes, i.e. Markov Chains (see Appendix A for a general statistical overview of these theories) enable one to consider also the dynamics of traffic that are observed in congested conditions, i.e. the effect of past conditions on the actual and future conditions. The convenience of this approach in comparison with the microscopic approach is in its faster computing times.

A probabilistic approach gives the opportunity to analyze the statistical properties of traffic and give estimates of the expected conditions and of the variability of traffic via the computation of e.g. standard deviation or 10-90% confidence values. Modeling the dynamic and the stochastic character of queues and delays at signals is needed for the following reasons:

- *Theoretical improvements and insight into the behavior of traffic performances*: traffic flow models should describe the transportation system as well as possible. The available models are bound to have rather simple formulations for reasons of tractability and they are often limited by some assumptions that simplify the state and control variables or their effect on the way vehicles propagate on the network. Therefore, improvements of these models still need to be performed both at the operational level (e.g. insight into the network capacity, throughput, clearance times, etc.) and at the behavioral level (e.g. car-following behavior, gap acceptance, etc.);

- *Assessment of existing management strategies*: the effects of existing DTM measures, like traffic control, need to be evaluated via improved models of travel times and their effects on the demand for traveling. In some studies it is also important to have knowledge of the confidence levels of these effects;

- *Design and planning of new infrastructures and control strategies*: insight on how one or another management strategy or intervention on the road infrastructure affect the transportation system reveals how a desired change in the system can be achieved. A deeper insight into the dynamics and the stochastic character of traffic through improved models can result in a more efficient set-up of strategies in time and degrees of confidence for achieving the desired results can be evaluated;

- *Effects of travel time on the network flows*: a better understanding of the dynamic and stochastic character of travel times at urban networks may improve the estimation of flows at urban networks, i.e. the route flows and their behavior in time. The knowledge of the variability of travel times can be used to evaluate the drivers' value of travel time uncertainty. These features can be for example useful in Dynamic Traffic Assignment (DTA) problems;

- *Improvement in short-term model-based travel time predictions*: the knowledge of how likely the traffic is going to perform in time and how large this information can be uncertain can be used in model-based travel time prediction and control problems.

This thesis shows that all available analytic models lack in describing theoretically the dynamic and the stochastic character of overflow queues and delays at signalized intersections, which is instead caught by the Markov model. A new expression for the expectation value of the overflow queue is derived from the data simulated by the Markov model, which is shown to fit well this dynamic behavior. Moreover, an expression for the standard deviation of the overflow queue is also proposed, which represents to the author's knowledge the first time-dependent expression for this measure that can be found in literature (an expression of the standard deviation of the delay was recently proposed by Fu (Fu 2000) under the assumption of stationary demand conditions for the whole period of analysis). The estimation power of the probabilistic models, together with their relatively simple formulations, motivates a much larger research in this direction, as it is shown in Chapter 8. Probabilistic formulations can therefore be done for more complex situations than the single lane, fixed controlled intersection; examples are given in the context of arterial corridors, multiple lanes and vehicle actuated controls.

1.2 Research objectives and scope

This thesis aims at giving a thorough analysis of the effects of traffic dynamics and travel time variability and to provide tools for improving estimation and prediction of urban travel times. The following objectives have been pursued in this study:

- To improve dynamic and probabilistic modeling of traffic flows at controlled intersections;

- To gain insight into the travel time variations caused by the variability of the demand and supply systems;

- To develop a model that is flexible and general enough to model the behavior of traffic at more complex control areas, intersections with multiple lanes and traffic streams, with different types of control systems, area network controls, etc.;
- To derive a formula for the time-dependent expected value of the overflow queue length in time, which improves the analytic delay estimation by considering the stochastic effects in time of congestion;
- To derive a formula for the standard deviation of the queue length, which may improve planning and design problems that aim to estimate the reliability of a transportation network; this measure can be also helpful in the estimation of flows if the users' choice process considers explicitly a cost of travel time uncertainty.

The description of this research is limited to motorized vehicles, in particular to passenger cars, while no attention is given to different vehicle classes and to the effect of one specific class on the others. Therefore heterogeneity of the traffic composition is not covered in this thesis. Moreover this thesis refers particularly to the delay incurred by vehicles at signalized intersections, therefore non-signalized intersections, roundabouts and uninterrupted facilities are not explicitly considered. The variability of traffic is intended to be derived from both within day and day-to-day variations, although the assumption of a known probability distribution may be different from the one observed during a day or at different days; therefore, refinement of the assumed probability distributions may be deduced from direct observation of traffic (i.e. by differentiating the day of the week, or peak hours and off-peak hours etc.). Chapter 2 discusses this issue in more detail.

1.3 Thesis contributions

1.3.1 Scientific contributions to the state-of-the art

This thesis contributes to the state-of-the-art of traffic flow modeling at urban signalized intersections in various ways:

1. It gives an empirical analysis of the relationship between the variability of the demand and the variability of travel times at urban networks (Chapter 2). Knowing the variability of travel times is shown to be as important as knowing their expectation value both for the traffic analyst and for the road traveler.
2. An exact probabilistic formulation of the queuing process within a cycle has been developed in Chapter 4. This model, combined with the cycle-by-cycle Markov Chain queuing process used already by other authors in the past (e.g (Van Zuylen

1985), (Olszewski 1990), (Brilon 1990)) enables one to estimate the probability distribution of queues and delays dynamically and for general arrival patterns.

3. Behavior of the expectation value and of the standard deviation and their mutual relationship has been analyzed, uncovering the underestimation error that one makes by neglecting the dynamic effects created by the random nature of traffic.

4. In order to obtain an analytic expression of the queuing process over time, which overcomes the elaborate computations required by the Markov model, a new formula for the expectation value of the overflow queue length has been derived (Chapter 6, referred to as the *Van Zuylen-Viti model*). This time-dependent model improves the available analytic expressions in that it considers the effect of the variability of the traffic states (demand, capacity etc.) under the following assumptions and properties:

    - It considers the stochastic effects when queues are both increasing and decreasing. No such effect was modeled explicitly so far also when queues are decreasing. If long queues need to be cleared and the signal operates often near capacity these effects can be very important and expected clearance times are considerably longer if for example they are estimated by a deterministic model.

    - It models the expectation value of the overflow queuing process also for non-stationary demand conditions, allowing one to model this process dynamically and simulate the transition between congested and uncongested conditions and vice versa, e.g. in peak period analyses.

    - It models the dynamics of the expectation value as a continuous function, which can be a desirable property in e.g. optimization algorithms. The first derivative is step-wise continuous if one assumes a step-wise demand to model non-stationary conditions, e.g. in Dynamic Traffic Assignment problems.

5. In order to have an estimate of the variability of overflow queues a new approximate expression of the standard deviation is provided in Chapter 6. The model has a similar formulation to the one of the expectation value and it shares the same properties and assumptions described above.

6. Consistency between macro, meso- and micro models in estimating traffic for long term planning and design problems is established at isolated intersections. The Van Zuylen-Viti models of the expectation value and the standard deviation of the overflow queue length are shown to give statistically the same results as the Markov model and the results of widely used commercial microscopic simulation software (Chapter 7).

7. The probabilistic approach is shown to be suited for more complex scenarios than the isolated, single lane fixed, controlled. Examples of these models are given in the context of arterial corridors, multiple lanes and vehicle actuated signals.

The probabilistic models are recommended for various applications in the transportation practice (Chapter 9). Planning, design and operational problems will improve their estimation results if an improved model of travel times is applied.

1.3.2 Research relevance and practical contributions

This research is relevant for planning and design problems, which involve a cyclic service process, especially when large random fluctuations of the state variables (demand, capacity etc.) are observed. This is the case of signalized intersections as well as ramp meters, toll plazas, etc. It can also be used to give a probabilistic description of delays due to e.g. incidents.

The Markov Chain process presented in Chapter 4 represents a very powerful technique for modeling such processes, since it treats variables at the probabilistic level and it simulates traffic by exact expressions based on mass-balance equations. This computing property makes mesoscopic models more suitable than microscopic models for planning purposes, since they simulate the variability of traffic and analyze causes-effect relationships between state variables, control variables and performance measures within reasonable computing times. Moreover, Chapter 7 will demonstrate that mesoscopic models give results consistent with microscopic models that are simulated under the same assumptions. On the other hand, modeling traffic operations with mesoscopic models enables one to obtain more accurate results than analytic models, which are based on more limiting assumptions, especially in dynamic networks. Analytic models and probabilistic models share the desirable property of directly relating the state and control variables to the performance measures (e.g. easiness in calibration and model validation), but they are capable of giving better insight into the statistical properties of these relationships. The models derived throughout this thesis using a probabilistic approach (Chapters 4, 6, and 8) for the signalized intersection under different inflow (stationary, non-stationary), geometric (single lane, multilane, isolated or in a network), control set-up conditions (fixed, pre-phased, dynamic control) represent only a few examples of application of this technique.

The property of Markov Chains to generate data within little computation times allows the traffic researcher to analyze the sensitivity of performance measures from the state and control variables and to derive simple heuristic and easy-to-use models when no simple formulas can be obtained by using exact macroscopic relationships. One example is given by the Van Zuylen-Viti models for the expectation value and the standard deviation proposed in Chapter 6. Applications in many other traffic problems and contexts are expected to give large contributions to a better insight on how to dynamically model traffic.

The analytic queue models presented in chapter 6 have particular relevance for traffic practitioners, since they give quick estimates of expected conditions and the possible uncertainty associated with these estimates in planning and design problems. On the practical importance of analytic functions for simulating travel times for design and planning purposes one may refer to Rose et al. (Rose 1989). These models are also valuable tools to apply in model-based travel time estimations and predictions for travel information systems and to analyze the effects of travel time on the distribution of flows along a network.

This thesis gives also valuable insight into the variability of queues and delays a traveler can experience driving through a signalized network. This information can be used by traffic analysts to evaluate reliability measures like travel time reliability, capacity reliability, network reliability etcetera (see chapter 2 for the definition of these measures).

1.4 Thesis outline

This section briefly describes the contents of each chapter of this thesis and the connection between them.

Chapter 2 gives an empirical overview of the causes and the effects of travel time variability at urban networks. The relationship between this variability and the day-to-day and the within-day dynamics of the travel demand is analyzed by looking at real traffic measurements. This variability is shown to highly affect the level of service of a network; a probabilistic expression of this level of service (in time) can be therefore derived by knowing the probability of travel times to be experienced by the driver. An estimate of the travel time variability is also shown to be fundamental in reliability studies and in flow estimation methods (e.g. DTA), which have an explicit driver's cost for uncertainty.

Chapter 3 provides a state-of-the art of analytic queue and delay models. It discusses the modeling implications of choosing this approach instead of a microscopic approach. Theoretical and approximate approaches have been described for both steady state and dynamic conditions, as well as for isolated intersections and arterial corridors, and for fixed-timed and time-dependent control schemes. These models and their limitation due to their simplifying assumptions will be discussed. In particular, all models were found to be deficient in dealing with the dynamic and stochastic behavior of queues when the signal operates near capacity. As a consequence, these models are not very well suited to analyze dynamic situations, e.g. peak hours. Moreover, very little is known about the variability of these queues, since very few studies analyzed this issue.

Chapter 4 describes the probabilistic approach to the modeling of overflow queues, which considers the process as a one-step Markov Chain process, i.e. the probability distribution of the queue after one cycle depends only on the distribution at the previous cycle and the arrivals and departures during the cycle. This method allows the analyst to estimate and predict the dynamic evolution of queues and the propagation of their distribution in time, quantifying the uncertainty around this estimation and/or prediction. However, this model, does not tell anything about the dynamics of the queue length and the delay within the cycle. Therefore, a new formulation for the expectation value of the queue length and the delay within a cycle is presented. This model contributes to the state-of-the-art presented in that it is an exact formulation which enables one to consider the effect in time of the variability of arrivals within the cycle. The new formulation for the within-cycle queuing process was recently proposed (Van Zuylen 2006)

Chapter 5 analyzes the statistical properties of overflow queues and delays using the Markov model. Analysis of the different behavior that can be observed under these assumptions has revealed that conditions of traffic in the neighborhood of saturation are strongly influenced by the random nature of demand and supply systems, creating an overflow queue and delay, which can be much larger than uniform and incremental delay components. Combined analysis of average and standard deviation of the queue in time shows strong interdependence among these two characteristics, especially in saturated conditions of traffic, therefore the ratio between standard deviation and mean influences the dynamic behavior of queues. This implies that an analytical expression for the standard deviation is also an important research issue. This chapter is inspired by earlier works of the author ((Van Zuylen 2003), (Viti 2004) among others).

Chapter 6 provides a new overflow queuing formula for the expectation value, which can be applied in the delay functions for planning purposes described in Chapter 3. Using the data simulated with the Markov chain process as benchmark for the development of empirical models, heuristic functions have been derived for the expectation value and the standard deviation of the overflow queue length in time. These models have a broader area of use than official manuals as for example the Highway Capacity Manual 2000 (TRB 2000), since they reproduce the expected evolution of queues and their variability as a function of time, without the necessity to fix an evaluation period but they provide estimates for every cycle. The derivation of the formula was initially presented in (Van Zuylen 2003) and the parameters were estimated in (Viti 2004). The simplified formula was finally presented in (Viti 2005).

Chapter 7 compares the results of the probabilistic model with several microscopic simulations of a commercial software package, VISSIM (PTV 2003). For this comparison microsimulation represents the only valid alternative to field data since it is rather unlikely that one can observe in real life sufficiently long periods of stationary demand conditions. The consistency between the three approaches in various conditions of traffic validates the two lower-level methods. This represents an important contribution to traffic managers and practitioners, since it proves that the dynamics of the overflow queue are well estimated with all three different level-of-detail models. The consistency between the models also in a dynamic scenario with non-stationary demand rates implies that the Van Zuylen-Viti model may contribute to the development of improved network loading models in DTA Processes. This chapter is based on a benchmarking study made by the author (Viti 2006).

Chapter 8 proposes the application of the Markov approach described in chapter 4 in three directions: arterial corridors, multilane sections and time-dependent controls. While there is very little difference in the formulation of the Markov model for isolated intersections from an intersection within a network, i.e. the shape of the arrival distribution, modeling the interactions between lane choice of drivers and queue lengths appears more complex. To account for this interaction the Markov model has been combined with a probabilistic lane-changing model. By doing so, the distribution of arrivals has been shown to have a dynamic character, according to the dynamic character of the overflow queue length. Furthermore, the Markov model at multilane sections allows one to account for spillback effects, which is useful information for a correct estimation of delays and for the design of exclusive turning lanes. Finally the assumption of fixed control settings has been relaxed by formulating a probabilistic model of vehicle actuated controls. This approach allows one to compute the probability of green time extension depending on the variability of arrivals and their headway distribution in time. The probability of overflow queues is computed accordingly. These model extensions are inspired by recently presented works ((Viti 2005), (Viti 2005), (Viti 2006)).

Chapter 9 discusses the potential applications of the Markov model and the potential future developments that can be expected with this modeling approach. Examples of application have been given for design and planning problems, in the modeling of traffic operations, in the travel time estimation and prediction problems etc.

Chapter 10 concludes this thesis and gives future directions of research.

The following flowchart, drawn in Figure 1.1, explains the relationship between the next chapters of this thesis.

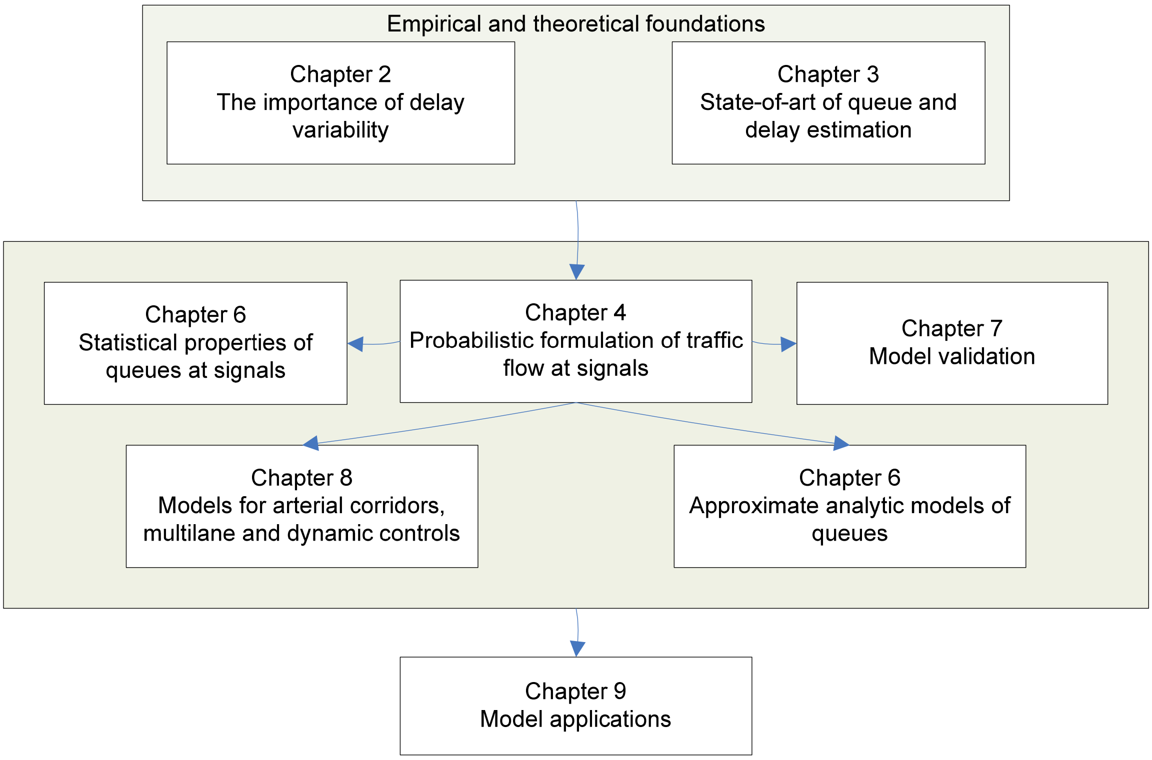

Figure 1.1: Structure of the thesis chapters

2

Causes and effects of travel time variability in urban networks

2.1 Introduction

The advancement of technology and informatics applied to the transportation systems gives opportunity for a more efficient use of the road infrastructure. Nowadays, transportation policies in large cities and metropolitan areas are giving increasing attention to the development of dynamic strategies of traffic management designed to reduce congestion by using these new technologies. Among these, Dynamic Traffic Management (DTM) strategies are designed for an optimal use of the network, achieved by adapting dynamically the road capacity to the demand (i.e. dynamic traffic control, speed control, ramp metering etc.) or by redistributing the demand in time and among all routes according to the available network capacity (i.e. in-vehicle or en-route information and guidance systems).

Travel times have important role in the assessment of DTM measures, since they affect the travelers' choices and they are determinants of the attractiveness of network links or routes. A good estimation or prediction of network flows together with the corresponding costs as a function of the applied DTM strategies has therefore a central role in the optimal set-up of such measures. However, estimation and prediction of travel times is largely affected by the complex structure of urban networks and by the dynamic and stochastic behavior of demand and supply systems.

The development of management strategies designed to adapt the road capacity to the actual demand (e.g. responsive controls) or vice versa (e.g. in-vehicle guidance systems) have supported an increased research over the causes and the effects produced by the variability of traffic. A quantitative definition of the role of the variability of travel times may improve the assessment of these DTM measures, since it can explicitly control the effects of these measures on the dynamics and the stochastic behavior of the transportation system. The knowledge of the variability of travel times is valuable for example to quantify the reliability and robustness of a transportation system and to assess the impact of traffic responsive and adaptive control systems. A deeper insight into what causes this variability can tell what part of it is systematic (or recurrent) and therefore predictable or controllable.

2.1.1 What causes travel time variations?

Travel time variability stems from several reasons and its relationship with the demand and supply characteristics has not yet been clearly defined. This problem is especially challenging in congested networks, since travel time variability increases with road occupancy and congestion (Van Lint 2004). The stochastic nature of the demand is widely acknowledged to be one cause of travel time variability (see e.g. (Clark 2005)). Variation in road capacity is also shown to strongly affect travel time (un)reliability, especially on motorways (see e.g. (Tu 2006)). Debate is nowadays around which of these two variations causes the largest variability of travel times.

Day-to-day as well as within-day demand variations are in fact observed in all transportation networks. The way activities are scheduled and their distribution in an area can give a rough estimate of how demand can be distributed in time and space. Large variations can be observed during the day and among days and in some cases the network is not able to serve the demand at a certain time and queues and delays are frequently observed. In other times of the day these parts of the network can have a large part of the capacity that remains unused. The network use is thus unbalanced during the different times of the day. These large variations of the traffic states have a degree of predictability, which can be improved by selecting an appropriate model (e.g. modeling travel times by specifying the time of the day, the day of the week and even the season as determinants of this variability). There are also variations in the traffic demand, which cannot be explained by specifying external factors. Part of travel time variability remains unexplained as it stems from the random nature of human behavior, i.e. their driving behavior and their travel choices.

Travel times at urban networks are for a large part determined by the delay drivers experience at (controlled) intersections; the variability of control delays is, in these systems, mainly affected by the variations of the demand, although the variations of capacity can also be large (e.g. in case of priority rules for public transport modes). The important role of the stochastic character of the demand is proved by the existence of several models, which explicitly assume stochastic distributions for the arrivals at the intersections. This stochastic character justifies for example the existence of a random delay component directly related to the variability of the demand in most of the delay formulas developed since the seminal work of Webster (Webster 1958). This component has very little role at low demand rates, while it represents the main component of the average control delay if the intersection operates near capacity. The demand variations at signal controls are therefore affecting travel time variability especially in their relation with the capacity. The importance of the relationship between demand variations and signal capacity has supported the research throughout this thesis.

This chapter gives insight into what causes travel time variability; variability of travel times is discussed both from the point of view of the traffic analyst and of the road users (Section 2.2); in particular a differentiation between what is predictable and what remains uncertain is discussed (Section 2.3). The value of uncertainty in the departure time-route choice process of travelers is later discussed using a survey (for more details see Appendix B) to give an idea of how people value uncertainty with respect to their value of time in their learning process and how they combine their experiences with information (Section 2.4). The different viewpoints of looking at travel time variability motivates the various definitions of reliability, which can be found in literature (Section 2.5). Finally Section 2.6 gives a synthesis of this chapter and summarizes the motivation for the research presented in the remaining parts of this thesis.

2.2 Predictable vs. uncertain

Travelers estimate and predict the costs of their trips by experiencing individual travel times. Past travels enable the drivers to build their personal experience, perception and opinion on an alternative for traveling. Past experiences can therefore tell the driver something about the uncertainty of these costs too. The traffic analyst does not know in general each traveler’s individual travel time but only aggregated travel times can be observed. Some monitoring tools (e.g. loop detectors) enable the road manager to estimate and predict average travel times. Some other monitoring techniques can trace the single trajectories (e.g. cameras, number plates recordings etc.) allowing the manager to collect individual travel times. Both with the collection of average travel times, for example during several days, and with the collection of individual travel times, the road manager can estimate the variability of these characteristics.

The notion of travel time variability is closely associated to the concept of (un)reliability. By its nature, reliability implies a notion of repetition and regularity (Bates 2001). Transportation networks are affected by various sources of uncertainty, which influence the drivers' choices. The sources of uncertainty stem from demand and supply variation. The variation in demand can be recurrent and show some cyclic properties (like daily or weekly traffic patterns), or non-recurrent (e.g. events like strikes or football matches), or by the variation of travelers' behaviors. The variation of the supply system may be a temporary or permanent (e.g. road works, construction of a new road etc.) and caused by external factors like adverse weather conditions, incidents or natural disasters. A good prediction of the expected conditions by accounting for day-to-day and within day traffic dynamics can catch partly this regularity, but uncertainty due to random unpredicted fluctuations will still affect this prediction. Since unpredicted variations can occur because of demand or supply variations, and they can be reduced by a change in both characteristics, many concepts of reliability have been proposed. A classification of these measures is given in Section 2.5.

From the travelers' viewpoint, the evaluation and prediction of route travel times and their variability is done by means of individual experiences during past trips. It is still not completely clear how people build their own opinions about each alternative of travel, and especially how much they value travel time variability in their choice process. Past studies have proposed to include the travel time uncertainty for example by considering the predicted travel time as sum of the average experienced travel time and an extra time as safety margin related to the uncertainty felt for that trip (e.g. (Uchida 1993), (Luo 2003)). These approaches have been on the other hand considered too simplistic to encompass the discrepancy between choice model predictions and observed travel choices. Different risk attitudes, memory skills, taste for habit and curiosity etcetera influence the perception of travelers and their learning process towards the most convenient choice. The value of reliability for the travelers and their decision-making process under uncertainty are not central issues in this thesis, although section 2.4 and appendix B briefly discuss the effects of travel time variability in the route-departure time choice process. For a more detailed research on these topics one can refer to e.g. (Van Berkum 1992), (Bates 2001), (De Palma 2005), or (Bogers forthcoming).

The concept of reliability used in this thesis is strictly connected to the uncertainty and the variability of travel times in the network, and in order to define the network reliability one needs to understand how the characteristics of the transportation system determine uncertainties and how these uncertainties influence the travelers' behavior.

2.3 Determinants of travel time variability

The following Figure 2.1 schematizes the various factors which determine travel time variability in a general network. This scheme does not consider the eventual re-routing or departure time adjustments due to variations in travel times and delays.

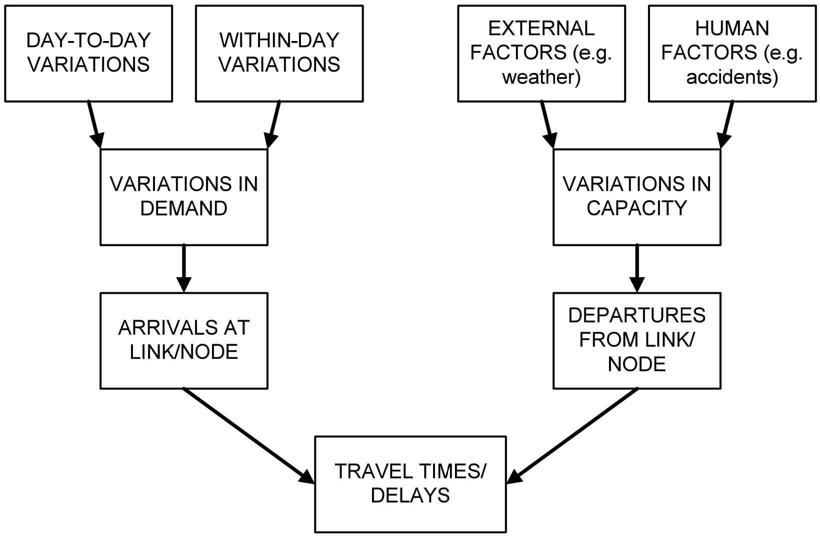

Figure 2.1: Factors influencing travel time and delay variability

The following of this section describes the role of these factors to the dynamic and the stochastic behavior of travel times.

2.3.1 *Day-to-day demand fluctuations*

The demand for traveling is primarily governed by the way activities are scheduled. For example, strong differences in flow rates are measured between working and non-working days, or between summer and winter seasons, both because of the concentration of holidays during warm seasons and because of an increasing use of the vehicles during cold seasons instead of other modes for traveling (bikes, motorbikes, etc.).

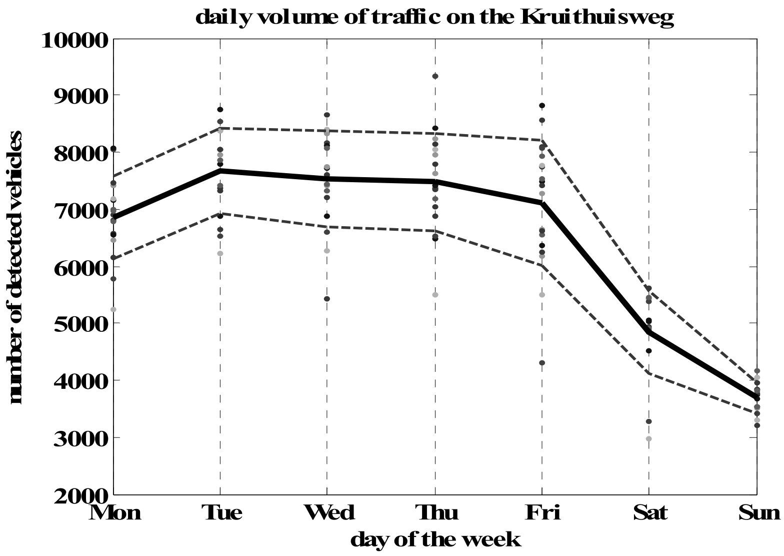

Figure 2.2: Total daily volumes of traffic detected on the Kruithuisweg (Delft) during October-December 2000 with resultant average and confidence intervals

Figure 2.2 shows how daily volumes of traffic changed in the autumn of 2004 in an urban road of Delft, in the Netherlands. The total number of vehicles within a day was measured with loop detectors placed under the road surface. It looks that the average daily demand does not change consistently during working days (Monday to Friday) while it is considerably lower during Saturdays and Sundays. Analysis of the standard deviation shows a relatively small variation of daily traffic within each weekday (around 10% of the mean).

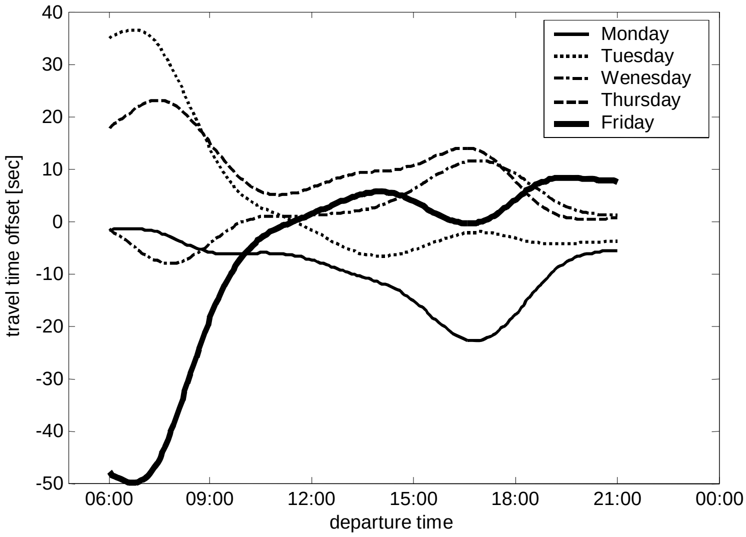

Figure 2.3: Day-of-the-week travel time offsets estimated on a Dutch motorway (Van Der Zijpp 2002)

Some studies (e.g. (Van Der Zijpp 2002), (Thomas 2006)) evaluate the predictability of travel times on motorways and urban roads by specifying the impact of daily, weekly, seasonal and weather variations. Van Der Zijpp et al. evaluated this impact by considering travel times as sum of a “typical” travel time, calculated using a whole-year data, and travel time offsets, which are determined by these variations. Large improvements in travel time predictions are achieved by simply gathering and analyzing historical data from a location to forecast the expected travel time and its variability.

Figures 2.3-2.4 show the estimated daily variations (represented by time offsets from the average) observed in a whole year analysis on a Dutch motorway (A13) in the year 2000. Figure 2.3 shows the offset one should take into account at each day of the week as compared to the average daily pattern; Figure 2.4 shows instead the same variations if one specifies which month.

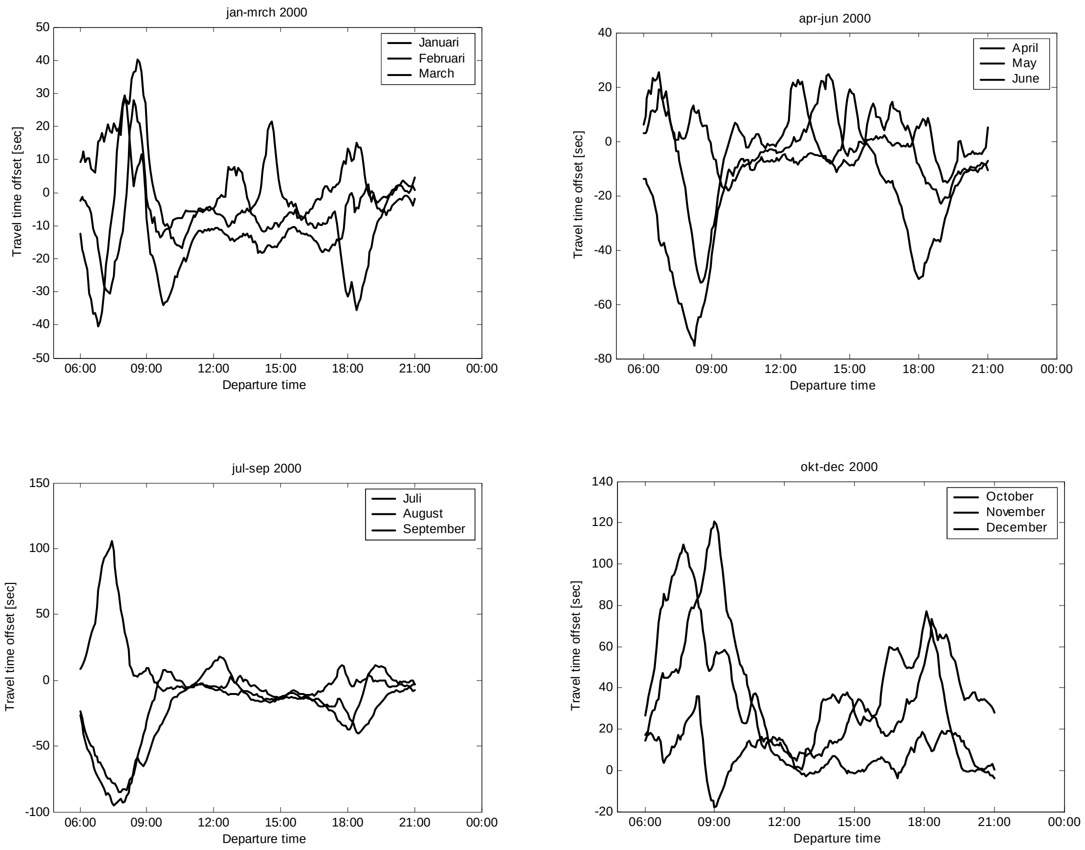

Figure 2.4: Monthly variations as compared to the average working day (Van Der Zijpp 2002)

Day-to-day variation can be caused also by the property of flows to adapt themselves to changes in the system, for example if new roads are built, or the capacity of some existing roads is changed (i.e. due to road works), or new traffic policies are adopted. These changes are certainly rarely predictable using historical data and only model-based approaches can give an expectation of the future traffic conditions (e.g. demand forecasting models, traffic assignment processes etc.).

2.3.2 Within-day demand fluctuations

The way activities are located and scheduled, and the limited network capacity, determines the existence of peak hour congestion in most of the big cities. Fluctuations of the demand pattern during a day are therefore determined by the desire of road users to travel during their most convenient time and route to reach their destination. If travel demand is low, there is little interaction among vehicles and the travel time is nearly constant, equaling the so called *free-flow travel time.* If travel demand is higher, but less then the capacity, vehicle interactions force drivers to reduce speed, leading to a slight increase of travel time. When the road space is insufficient to serve all vehicles driving on one location at the same time some travelers experience uncomfortable driving situations; the interaction between drivers is very strong and the behavior of one vehicle is conditioned by the behavior of neighboring vehicles and stop-and-go and queuing phenomena are frequently observed. Some drivers anticipate or postpone their time of departure or change route to avoid these bad conditions but sometimes this adapting behavior is not sufficient to reduce the demand below the actual capacity.

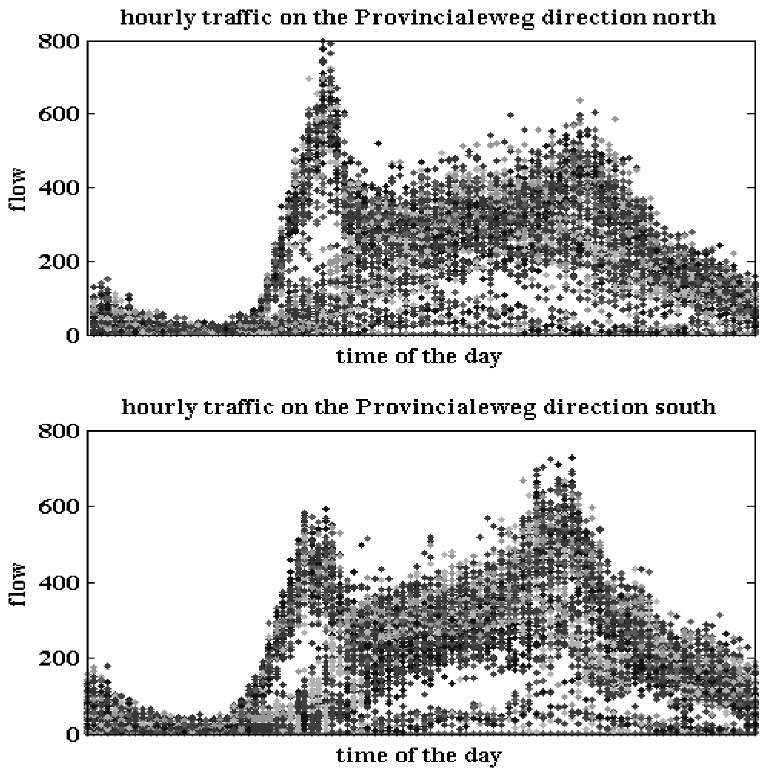

Figure 2.5: Hourly volume of traffic detected on the Provincialeweg (Delft) for both directions

Figure 2.5 shows how traffic can be scattered. The variation of the demand in time is much more evident than in the total daily volume. Moreover, the figure clearly shows the influence of the activity locations. In fact, the highest peak in direction north appears in the morning, while in direction south the largest volume is observed in the afternoon.

Since starting time and location of many activities are concentrated in certain points in time and space, peak periods frequently lead to congested conditions, increasing travel times on links. Figure 2.6 shows the effects of the peak period in terms of travel times (in seconds) again on one arterial road in Delft. The capacity of the road is sufficient to keep travel times at low values for most of the time. In the morning the travel time is much larger than in off-peak hour and the standard deviation is comparable with the average.

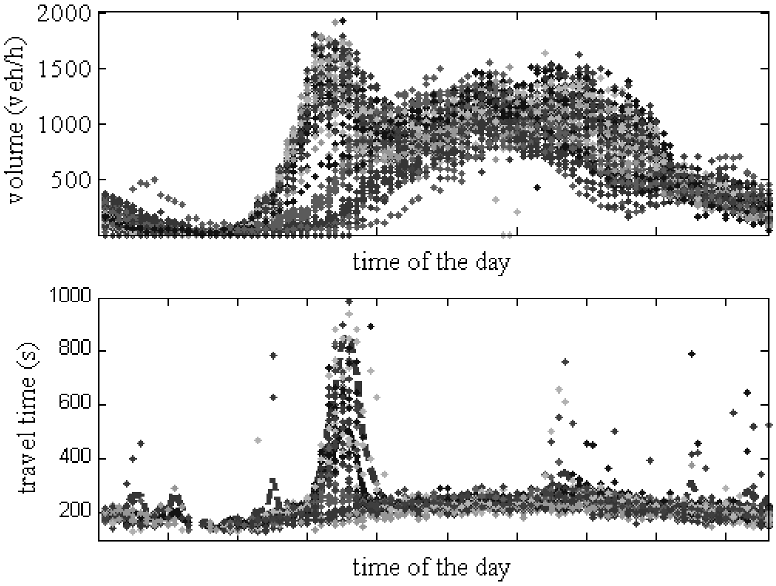

Figure 2.6: Hourly volume of traffic detected on the Kruithuisweg (Delft)

Fluctuation of the demand during the day makes road capacities inefficiently utilized. They are in fact observed in this case both for the time the peak is observed and for the duration of congested conditions.

2.3.3 Variations in capacity

Traffic composition, heterogeneous speeds and differences in travelers' driving behaviors in traffic networks are also causes of variable travel times. This issue is dealt with marginally in this thesis although the hypothesis of stochastic arrival distribution at the traffic signal assumed in the next chapters encompasses these variations.

Variability between vehicles that make the same journey in the same period (*inter-vehicle variability*) is very difficult to catch in a macroscopic viewpoint, while it represents a fundamental characteristic at the microscopic level, for example in modeling the variations in car-following behavior of travelers (Ossen 2006). Tampere (Tampere 2004) investigated the dynamics of congestion in conditions of flow near capacity by assuming some source of variability in vehicles or drivers' characteristics. This variability can be both in traffic composition and drivers' behavior (e.g. different desired speed, or car-following and lane changing behavior etc.). He concluded that variability in driving behavior increases the probability of breakdowns and affects traffic condition stability. This issue has supported the development of driver's assistance systems designed to reduce these perturbation effects and consequently the risk of breakdowns at conditions near congestion. Advanced vehicle guidance systems like the Advanced Driver Assistance Systems (ADAS) are therefore expected to reduce travel time variability by constraining drivers to more homogeneous drivers' operations, e.g. lane changing, gap acceptance and platooning (Minderhoud 1999).

At urban networks the variable traffic composition seems to have a larger impact than the speed variations and driving skills of the drivers, but up to date no empirical evidence is known. Some studies use microsimulation to investigate this issue (e.g. (Kang 2000)). Large decrease of capacities has been estimated at urban controlled intersections with an increasing presence of heavy vehicles and with priority rules for public transports.

A behavioral approach like Tampere's may partially explain the large variations in vehicle throughput, which can be observed in real life, but these variations do not fully explain why different capacity values can be measured. Reductions of capacity are also due to external factors, which implicitly modify both travelers' driving behavior and road physical capacity, i.e. degraded road pavement, bad weather conditions and visibility etc. A different approach than the one of Tampere can be found in Brilon et al. (Brilon 2005), where the authors do not analyze explicitly the causes and the effects of stochastic demand propagation on travel times, but they analyze directly the road capacity as a stochastic variable.

Apart from stochastic variation of capacities due to the random nature of the demand, other factors can influence the capacity and cause unexpected fluctuations of travel times. Unexpected or non-recurrent events can affect temporarily or permanently the road capacity, e.g. accidents can block one or more lanes of a freeway, creating a bottleneck for the incoming vehicles.

Some of the capacity fluctuations can finally be caused by exogenous factors (e.g. adverse weather conditions, natural disasters, bad visibility etc.). For example, strong evidence relates the weather conditions with the travel time estimations (Van Der Zijpp 2002). Figure 2.7 shows how travel time offsets have been computed on the Dutch motorway A13 by considering conditions of good weather, light rain and heavy rain (taken from (Van Der Zijpp 2002)). As one can notice very bad weather conditions cause congestion to build up earlier, apart from being more severe. A model which considers adverse weather conditions as causes of an offset of travel times from the average can more accurately estimate and predict such extra-delays.

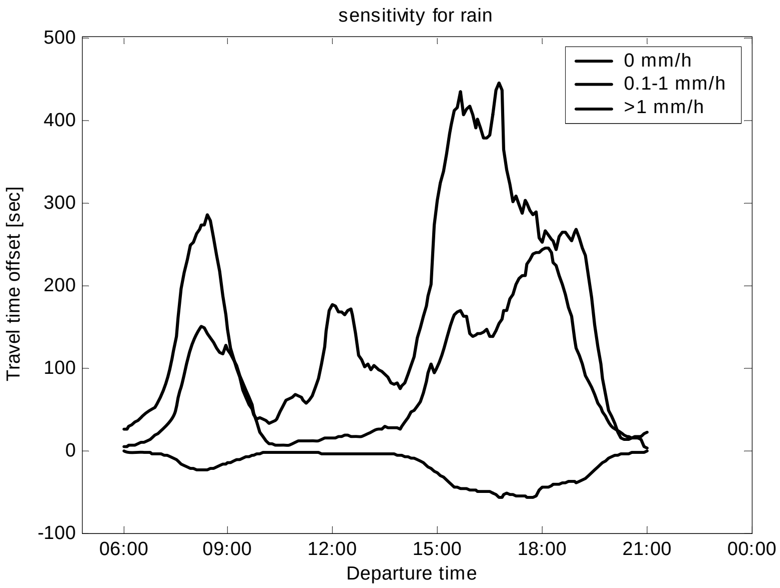

Figure 2.7: travel time offsets with different rain conditions

Combining this model with models that estimate delay offsets conditioned by departure time, day of the week and season as shown in Section 2.3.1, in which the trip is made, can further reduce the uncertainty on the travel time prediction (Van Der Zijpp 2002).

In conclusion, part of the daily travel time variability can be explained by recurrent (or periodic) variations. The predictability of these variations, from the point of view of the traffic manager, depends on the way the system is monitored and on the estimation model chosen to represent it. Travelers are also able to estimate in some way these periodic variations using their past experiences and by gathering information from any available source. Both managers and drivers cannot on the other hand achieve a perfect prediction of future costs because of random variations which are not driven by any explainable factor, and match between predicted travel time and actual travel time remains only probable. Knowing how variable travel times are expected to be can therefore quantify how much information can be reliable and trustworthy. Knowing how much travelers value this variability in their choice process is also valuable to understand how they will react on the available information and how much they will rely on their own perception or expectation of costs.

2.4 The value of uncertainty for the travelers

Assessing the influence of travel time variability on travelers' decisions has been addressed as one of the challenges in recent transportation research. Several authors emphasized the importance of including uncertainty about travel conditions as a factor influencing travel decisions, especially in terms of route choice and departure time choice (e.g. (Bates 2001), (Avineri 2003), (Chen 2004), (De Palma 2005)). The expectation of travel time variability from the travelers' viewpoint is strongly related to the concepts of perception and information. The greater the variation of these costs, the more difficult it will be for travelers to acquire reliable information and to perceive a correct expectation of the travel costs. Clearly, the ability to predict variations in demand will also vary within the traveling population (Bates 2001).

Variability of costs, conflicting objectives, competing alternatives and heterogeneous risk attitudes among drivers make decision making variable among travelers and consequently difficult to predict. The variability of travel times, which a driver or different drivers can experience, together with the variability of each driver's characteristics makes the prediction of their choices a challenging task. Some behavioral models assume travelers to choose their preferred routes according to the costs they expect for all known alternatives; in reality decisions may be made under complete or partial information about the real travel costs, or under time constraints, or even depending on the traveler's emotional state at the time of deciding (Bechara 2005).

Travelers get to know about how variable their next travel time can be from two sources: from their own past experiences, thus how many times they traveled using an alternative or another, and from the information they acquire from external sources, i.e. variable message signs, road maps, radio, internet etc. Past experiences and information are combined in order to have a higher degree of confidence on the expected travel costs (Bogers 2005). Travel times experienced and traveler's attitudes at each trip (his level of habit, curiosity, risk acceptance, experience and reliability towards roads and information) are combined to get an expectation of the utility the traveler might have from the next trip, and how uncertain these costs can be. In some situations travelers seem to prefer a reduction in variability than in the mean travel time ((Bates 2001), (Bogers 2005)).

2.4.1 Experience and learning mechanisms in car traveling

People get to know about the possible costs they may face by experiencing the routes for various days, and sometimes during different periods of the day. If a traveler has little knowledge about the status of the roads at the time of departure, he might assign a very large safety margin to the expected travel time, in order to avoid long travel times causing also very long delays at the arrival point. If instead the traveler has experienced the roads quite often, given his past experiences he may exclude the possibility that at the time of departure delays may occur. The level of experience may thus considerably influence the perception of uncertainty and the choice under risk.

The level of perception given by the experience is then an influencing factor in the travelers' decision-making process. This level of confidence may be influenced by the number of travel conditions experienced using an alternative, related to the variability of travel times, but also on travelers' individual characteristics like memory skills, habit, curiosity, stress etc. Travel time variability and experience are therefore highly correlated and this relationship can affect the way flows distribute along the competing alternatives. Based on former route and departure time choices, the travelers have personal experiences. From these experiences, they can learn about the characteristics of the routes they have chosen, about how to interpret travel information and about the reliability of this information.

2.4.2 Acquiring travel information

Experience is often combined with other sources of information (i.e. in-vehicle, en-route and off-route information systems) in order to get a better perception of which costs a traveler will experience until arrival at destination. While past experiences might help the traveler at having a guess of what kind of conditions are likely to be found on a certain route and time and how much a trip "usually" costs, information can provide a measure of the actual (or forecasted) status of the roads.

Information systems are therefore designed with the scope of directing the traveler towards the most convenient choice of travel. This effect is on the other hand related to various factors, which may cause the actual effects to deviate from the desired. Rerouting effects may simply move congestion from one location to another; information can have poor impact on the demand or be inaccurate.

Travel time variability affects in fact the quality of information and the impact of the latter to the users. Some users may still rely on their personal opinion instead of following the suggestions eventually given by the information if information is frequently imprecise. For the assessment of ATIS systems it is important to know how travelers combine past experiences with past information. Bogers et al. (Bogers 2006) give evidence that past experiences have less influence than travel information in an uncertain environment.

2.4.3 Individual characteristics

Even if travelers increase their level of experience and information in order to reduce as much as possible their uncertainty about the expected costs of a trip, uncertainty still affects their choice if travel times are variable.

Research on decision-making under risk and uncertainty has its origins in behavioral science like economy and psychology with the works of Bernoulli (Bernoulli 1738), who found in his studies discrepancy between objective variability and subjective perception of this variability. In the last two centuries this research has been carried on from different disciplines, economics and psychology among others, to define the human factors, which primarily influence this choice process. Travelers decide which alternative of travel to take accepting the risks that this choice implies. Even the with the same information and experience, people can decide to use different alternatives of travel for their next trip because of different risk attitudes. The strong relationship between risk and uncertainty motivated several studies from Bernoulli, which gave the foundations to the Expected Utility Theory (Von Neumann 1944). Studies in the risk behavior of travelers have suggested the investigation of both heterogeneity of risk attitudes among travelers (e.g. (De Palma 2005)) and individual non-linear risk aversion depending on the distribution of costs (e.g. (Avineri E. 2003)). Bogers et al. (Bogers 2004) found from a survey that truck drivers show a linearly increasing risk aversion with the standard deviation of travel times.

Apart from travelers' risk attitude, several human factors make the choice of travelers variable. Individual characteristics like habit, curiosity or short memory limit human minds and impede the correct estimation of expected costs (Bogers 2005). Drivers may value in a different way their driving times, or delays, or waiting times at queues.

2.4.4 Effects of travel time uncertainty in travel choices: a survey

Perception levels vary among road travelers due to different experiences, information used and individual characteristics. This is reflected in variability in travel choices. Special attention is given in this section to the effects of travel time variability on the route and departure time choice process of travelers.

The valuation of travel time reliability can be done by means of descriptive models, i.e. using Random Utility Theory (RUT). Even if several criticisms have been underlined, especially on the assumption of traveler as rational utility optimizer, this method is still useful to model the impact of different traffic performances and variables on drivers' choices. Since travel choices are often discrete variables (e.g. route, mode) or approximated as discrete (e.g. time of departure) it is difficult to associate these choices to continuous variables like e.g. travel time, or travel time variability. Discrete choice models (Ben-Akiva 1985) are therefore very appealing since they model the travel choices through utility maximization criteria: the traveler selects the alternative, which gives him the highest utility among the chosen set of alternatives. Utility functions are represented by all investigated variables together with parameters, which quantify the relative importance of a change of a variable on the travelers' choice. These parameters are therefore calibrated from a sample of observed choices.

Figure 2.8: Screenshot of the route choice phase in the TSL

This section shows some results taken from two laboratory experiments made with the web-based tool Travel Simulator Laboratory (TSL, (Hoogendoorn)), designed at the Delft University of Technology (Delft, the Netherlands). A description of the tool and the set-up of the experiments are given in appendix B. The role of uncertainty in choice behavior is investigated by analyzing empirical results of the choice process of travelers under uncertain conditions in terms of route and departure time. The respondents asked to repeat 25 times a certain trip with a fixed origin and destination on the motorway network around the city of Amsterdam (see Figure 2.8) and to select at each round a time for starting the trip and, after receiving the information about the expected status of the roads, the route to use.

Travel times have been calculated in TSL using a stochastic simulation model, which computes travel times according to a normal distribution and assigns a random sample to each respondent out of this distribution. Information is given, whenever the panel is on, displaying the length of queue in kilometers. To simulate the stress of waiting in a queue a waiting time proportional to the total travel time spent a scaling factor was applied.

The first experiment presents the choice of two routes with similar expected travel times and variance while the second experiment involves two routes where the travel time of the first is highly variable while the second has a markedly higher expected travel time but very high reliability (see Table 2.1). For the first experiment data from 52 respondents were available; for the second one there were 63 respondents. It concerned highly educated people mainly from The Netherlands, Italy and Portugal, most of them being engineers. For a more detailed explanation about set-up of the scenarios and the design of the experiments see also (De Groot T. 2004).

The value of past experiences and information in their decision-making process was analyzed in a simple scenario involving two alternative routes and an interval of possible departure times. After the collection of some individual characteristics like age, gender, driving experience etc. the respondents were asked to give their preference regarding the following travel time attributes: arriving early, in-vehicle travel time, arriving late. During the experiment, a score was determined by the sum of the normalized weights for a route attribute times the number of minutes spent for that attribute. The respondents were told that their goal would be minimizing this score: let α_{ik}, β_{ik}, γ_{ik} be respectively the weight assigned for a minute lost for early arrival, driving time and late arrival at destination, and $t_{ik}^{early}(n)$, $t_{ik}^{driving}(n)$, $t_{ik}^{late}(n)$ the respective quantity of minutes that respondent i loses selecting an alternative (route and departure time) k during step n. The score is computed for each individual by:

$$\mathrm{I}_{ik}(n) = \alpha_{ik} \cdot t_{ik}^{early}(n) + \beta_{ik} \cdot t_{ik}^{driving}(n) + \gamma_{ik} \cdot t_{ik}^{late}(n) \tag{2.1}$$

Table 2.1: Characteristics (in minutes) of the routes in both experiments

		Experiment 1	Experiment 2
Route 1	Average	17.68	23.05
	Standard deviation	4.25	6.65
Route 2	Average	19.86	25.31
	Standard deviation	3.92	2.43

The respondents' utility function for choosing the alternative k has been assumed linear and having a mixed Logit model structure, following the relationship:

$$U_{ik}(n) = \boldsymbol{\beta} \cdot \mathbf{X}_{ik}(n) + ASC_{ik} + \varepsilon \tag{2.2}$$

The indexes i, k, n represent as the above Formula (2.1) respectively the alternative index, the respondent index and the time step. $\boldsymbol{\beta}$ and $\mathbf{X}_{ik}(n)$ are respectively the vector of free parameters and the vector of explanatory variables related to person i and alternative k, which are found significant in the choice of users. ASC_{ik} represents the alternative specific constant related to person i and route k while ε is the error term, which models some choice variability. The choice of a mixed Logit is justified by the time-dependent characteristic of the choices for each respondent and the strict correlation among these choices. For a detailed description of the model and the model results see Bogers et al. (Bogers 2005).

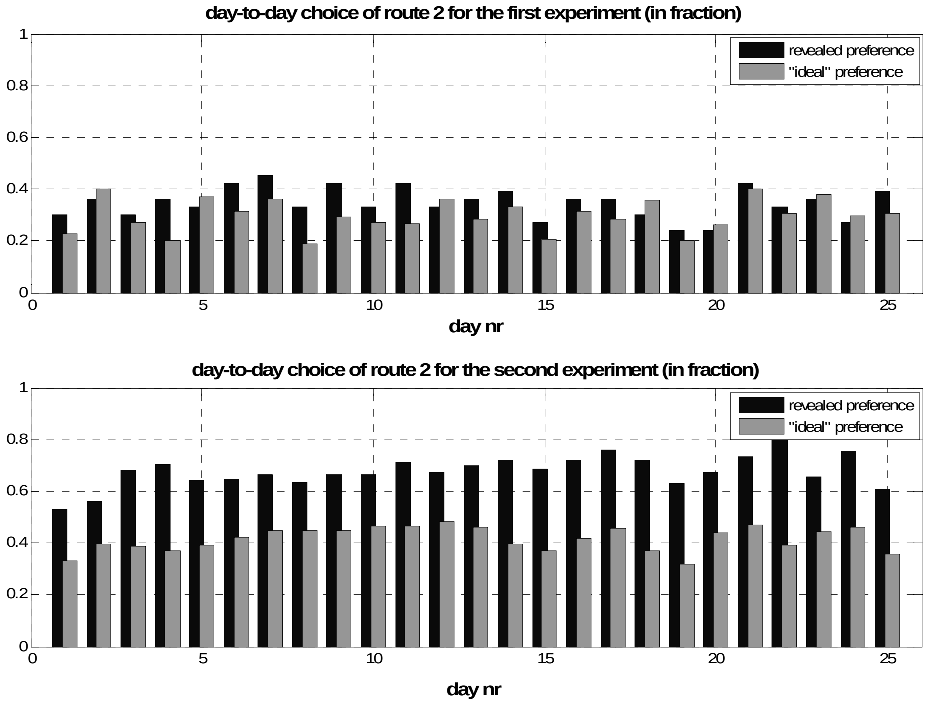

Figure 2.9: day-to-day fraction of choices for route 2 in the two experiments

Figure 2.9 shows the average preference of the respondents to the two routes in the two experiments in terms of fraction of choices. In the first experiment route 1 is generally preferred to route 2, according to the lower expected travel time. Average preference in fact does not exceed 40% during the whole experiment. Moreover, the fraction of people choosing route 2 is in accordance with the "ideal" preference, represented by the fraction of times route 2 had actually a better score than route 1. In the second experiment, route 2 is definitely preferred to route 1. Although the average travel time is higher, respondents preferred the reliable route on average 70% of the times. Comparing the revealed preferences to the "ideal" ones (the times route 2 was actually faster than route 1) there is a consistent difference, giving an idea of the importance of reliability in user's decisions and to what extent this characteristic may play a role in the utility of travelers.

The results shown in Figure 2.9 can point at two conclusions closely related to each other. The first is the relevance of the reliability of route 2 in the second experiment for the route choice of travelers. It appears that travelers on average overestimate the utility of the reliable route with respect to the actual expected costs, and are willing to refrain from saving some driving time in order to be more on time. Note that, during the preliminary steps, the majority of people declared to pay the most attention to scheduled delays, followed by driving time, and at last they care about arriving early. The choice of a faster route would have led to a better score, but respondents consistently chose the slower but more reliable route. In this choice process people show to be highly risk averse, as found also by Bogers et al. (Bogers 2004) in freight transport.

The second conclusion relates the penetration of the information given to the respondents to the variability of route 2 in the second experiment. In fact, looking again at Figure 2.9 it appears that respondents of the first experiment respond more thoroughly to the hint given by the information panel. This conclusion is confirmed by looking at the best-fitting values of the parameters in the mixed Logit models (see (Bogers 2005)). In both experiments the information given, represented as variable by the number of kilometers displayed by the road panel for each route, and the experience, represented by the last experienced travel time and the last scheduled delay, are found significant. In both cases the value of information is much higher than the experience.

In the first experiment the information given represents 80% of the utility, while in the second it represents only 45% of it. Even if the last travel time is a significant factor, it represents only around 2% in both experiments. Also the last experienced scheduled delay has little role in both experiments (respectively 8% and 3%). Experience is also controlled by a third factor, namely the number of times the route was previously chosen, which can be seen also as an indication of habit. This factor represents about 10% in the first case and even up to 27% in the second experiment. In this case travelers who have selected the reliable route are more likely to select it also in the future, giving less attention to what information they get. In conclusion, both the value of information and past experiences are affected by uncertainty.

Figure 2.10 compares average experienced travel time at each iteration to the average shortest ones (average of the smallest travel times at each simulation), displayed in the left two figures, and the average delay (here intended as both early and late arrivals according to their chosen departure time) to the delay people would have experienced if choosing always the fastest route (right two figures). Respondents in the first experiment have average travel times delays in accordance with the lowest scores, while in the second an increase of travel time (an average increase of 15%) is balanced by a consistent decrease of average delay, which is nearly zero. Looking again at Figure 2.10 one can have an insight of the learning mechanism of travelers. It appears that while the average travel time is not changing with the experience of the users in both experiments, the average delay clearly decreases in time.

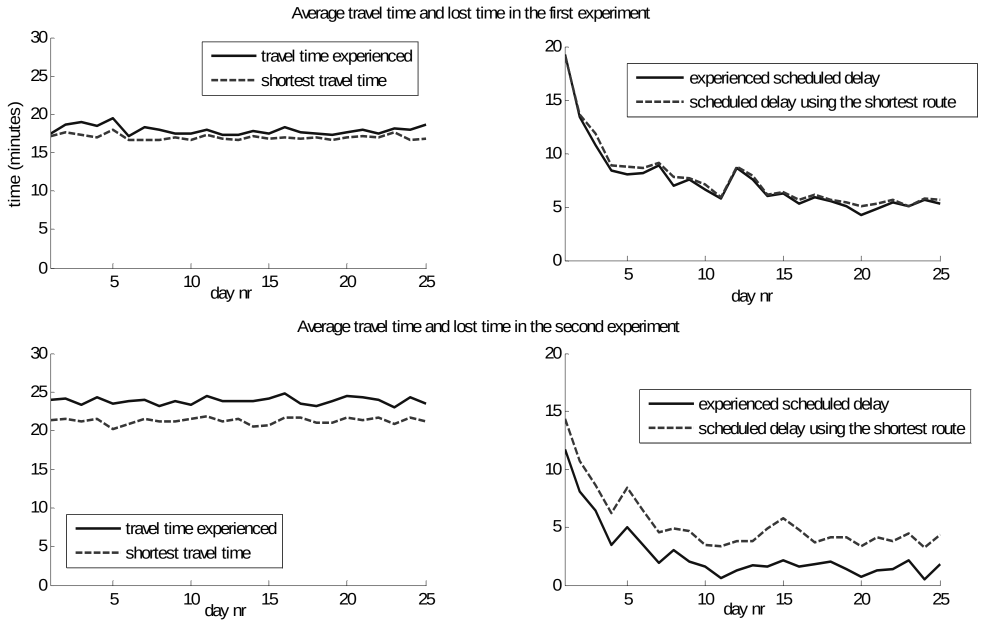

Figure 2.10: day-to-day evolution of travel times and delays in the two experiments

A conclusion that can be drawn from these results is that users are willing to spend more time on the road in order to guarantee a higher probability of on-time arrivals. The role of travel time variability is therefore important in the departure time choice of travelers too. This conclusion is supported also by Bates et al. (Bates 2001) who point at departure time choice to be the most sensitive choice level to travel time variability. In the first experiment the respondents chose their time of departure with a safety margin of around 11% with respect to the average travel time (arithmetic average of the two alternatives travel times) while in the second experiment this value is 6%. In this case it is interesting to note that the interval selected for traveling is nearly equal to the average travel time of the reliable route. The standard deviation of this interval does not appear to change consistently (around 0.5 of the mean in both cases).

The above experiments have shown empirical evidence of the relevant role of uncertainty and reliability of travel times in the decision-making of travelers. Both route choice and departure time choice are affected by uncertainty in various ways. As seen in the last section, reliability is valued in a positive way in the utility to an alternative of travel. This implies that the effort in estimating travel time variability should be comparable to the estimation of the average travel time. The various sources of uncertainty, which have been shown to affect the transportation system, yield different definitions of reliability, as discussed in the next section.

2.5 Measures of Reliability

Reliability and variability are often dealt with as related characteristics. Nevertheless, they are different in focus, they are measured in a different way, and they suggest different potential solutions. Lomax et al. (Lomax 2003) provides two distinct concepts for reliability and variability:

- **Reliability** is commonly used in reference to the level of consistency in transportation service for a mode, trip, route or corridor for a time period. Typically, reliability is viewed by travelers in relation to their experience. Travel time reliability is only one example but this concept can be related to other characteristics of the transportation system (e.g. link, network, etc.).

- **Variability** might be thought of as the amount of inconsistency in operating conditions. This definition takes more of a facility perspective and, therefore, relates the concern of transportation agencies. An example is the travel time variations described in Section 2.3.

The distinction in viewpoint underlined by Lomax et al. gives insight into the difference between the two concepts. Reliability is differently viewed if the individual traveler's perception or the average behavior of all travelers is analyzed. Moreover, different perception and knowledge of the variability is obtained from these two viewpoints, since the former is strongly dependent on the drivers' level of experience, while the latter on the way the system is analyzed and monitored.

Some objective measures that have relationship with reliability/variability are the following (Lomax 2003):

1) *Statistical Range*: these measures typically take the form of an average value plus or minus a value that encompasses the expectations for a certain % of the trips (e.g. the standard deviation encompasses about 70% of the trips, the double 95%). These usually appear as variability measures. Examples are the Travel Time Window (e.g. average travel time ± standard deviation), the Percent Variation, i.e. the ratio of standard deviation and travel time, or the Variability Index (ratio between 95% travel time during peak periods and 95% during off peak), which encompasses information regarding predictable fluctuations like day-to-day fluctuations.

2) *Buffer Time Measures*: these measures usually indicate the amount of extra time that must be allowed for a traveler to achieve destination in a high percentage of the trips. These measures illustrate reliability. Examples of this group are the Buffer Time, which gives the minutes of extra time needed to guarantee a statistically minimum number of arrivals within the preferred arrival time at destination, or the Buffer Time Index, which is the ratio between Buffer Time and average travel time.

3) *Tardy Trip Indicators*: These measures use a threshold value to identify an acceptable late arrival time. A measure of this class is the On-Time Arrival, which indicates the percentage of trip travel times that are within an arrival time window. The arrival time window is defined by the travelers' characteristics (e.g. importance of the activity at destination) and the expected duration of the trip. Another measure is the Misery Index, which weights the fraction of long trips by the average number of minutes lost with these trips.

In transportation reliability has also different definitions depending on which characteristic of the network this quantity is associated to. Clark and Watling (Clark 2005) clarify that the word reliability has meaning only if a performance measure is specified and it is measured in terms of probability only if a critical value is also specified. Here are some examples:

4) *Travel time reliability* is intended as the probability that a trip on a link or route can be completed within an upper bound assigned. Another definition can be related to the ratio between the travel time and a fixed percentile, e.g. its standard deviation. Another travel time reliability measure is the buffer time index (Lomax 2003), where the standard deviation in the ratio is replaced by the difference between 95th and expected travel time.

5) Since uncertainty in travel times depends also on the supply system and the way it connects all origins to the destinations, *capacity and connectivity reliability* represent also important factors, since they measure the probability that capacity is lower than expected.

6) In a DTM problem of how to provide traffic information to the users, the *reliability of information* represents a primal component. The more reliable the travel time estimated by the road authority, the more accurate information can be provided to the travelers. If information is very reliable, the perception of uncertainty (consequently the risk of doing non-optimal choices), may become very small.

The fundamental role of travel time variability and performance indicators like the standard deviation of travel times in all measures of reliability and variability motivates the research presented throughout this thesis, which focuses on the study of this characteristic at urban signalized intersections.

2.6 Summary

This chapter focused on the central role of travel time and travel time variability in transportation analysis, and especially to its characteristics of variability and reliability.

An overview of causes and effects of travel time dynamics and variability is given both conceptually and with empirical analyses.

Road managers who provide information to the traveler may increase the network performance by simply redirecting part of flows towards routes, which will give lowest travel times. To do so, the manager needs models able to give accurate estimates of expected travel times, as well as the uncertainty around such travel times. On the other hand, a road manager who plans a management strategy should also estimate well the travel times, queues, their dynamics, and the variability, which can affect the system, in order to effectively assess the effects of one or another possible control strategy.

Travel times at urban networks are characterized by their strong dynamic and stochastic character, principally because of the stochastic character of the demand. The next section will show that this characteristic is stronger when links operate near capacity. However, better estimation and prediction can be achieved with enhancing the road monitoring system and by using improved travel time models, only part of the demand fluctuations can be forecasted (e.g. day-to-day and within-day). This chapter has made a clear distinction between the various sources of variability that stem from the demand.

Travel time variations are dependent on the stochastic nature of the demand, but it is also vice versa. Travelers take in consideration travel time variability sometimes even more than mean travel times. Especially when time constraints are involved (e.g. important meetings) a road user is very concerned about the risk of long unexpected travel times and is willing to spend extra time on the road if this is more predictable and implies less risk of late arrival.

The importance given by road analyst and governments to the travel time reliability and the reduction of uncertainties, and the impact of travel time variability on road drivers should motivate as much research on these characteristics as for the expectation value of travel times. This has motivated the research presented in the next chapters. In particular, the research focuses on delays and queues at signalized intersections, which represent the main cause of travel time variations in urban networks as it is explained in the next chapter.

3

State-of-art of traffic flow modeling at signal controls

3.1 Introduction

Chapter 2 has introduced the notions of travel time variability, uncertainty and reliability in a general transportation context. The causes and the effects on travelers of these travel time characteristics have been described by empirical analysis of traffic data and by a survey. From these results, travel time variability has been shown to influence largely the travelers' choices, especially in terms of route and departure time choice. Since travel times at urban road sections operating near capacity are characterized by a large variability, a better insight into the dynamic and the stochastic behavior of these travel times is very important especially within this range. The following of this thesis refers in particular to delays at urban signalized intersections.

Urban travel time is usually subdivided into a travel time component spent for traversing links and one for passing the road junctions. Vehicles propagate in an urban network and modify their speed when interacting with each other. The delay experienced at one intersection depends, among other factors, upon the implemented control method (non-signalized or signalized control) and on the demand that needs to be served in time. This delay is determinant of the largest part of the total delay at urban networks, since non-signalized intersections are usually implemented only when low flow rates are involved, while roundabouts often require too much space. Moreover, waiting times at signals have a larger impact than running times in the perception of travel costs of most of travelers (Horowitz, 1978).

This chapter presents the state-of-the-art of queue and delay models at such intersections, with particular regard to the analytic models developed up to date. The next section defines the different components of route travel times in an urban context, defining the factors determining such travel times and their variability. Section 3.3 deals with the state-of-art and practice of analytic delay models at signalized intersections. The strong relationship between the queuing process and the dynamic and stochastic behavior of delays is clarified by the queuing theory described in Section 3.4 in the context of isolated, pre-timed controlled intersections. The assumptions and limitations of the presented methodologies are discussed in Section 3.5. Section 3.6 describes the modeling of delays at more complex control scenarios, i.e. the delay and queue models in the case of arterial corridors, multiple lane sections and time-dependent control schemes. Section 3.7 discusses the possibility to estimate travel times using a simulation approach. Finally Section 3.8 gives a synthesis of the chapter and gives the motivations for the methodology applied in the following chapters.

3.2 Traffic flow in urban networks

A schematic representation of an urban network is usually defined by a couple $(\mathbf{N},\mathbf{\Lambda})$ where $\mathbf{N}$ is the set of nodes and $\mathbf{\Lambda}$ is the set of links in the network. Each characteristic contributes to a part of the total network travel time. Link travel times are primarily determined by the time a vehicle takes to traverse the link, with a speed, which is also influenced by the presence of other vehicles on the road. Waiting times at nodes are determined by the interaction between different streams converging to the node and the way traffic is regulated.

A categorization of flows in a transportation facility is given in Kang (Kang, 2000):

- *Uninterrupted flows*: belonging to this category are all flows in road sections where the only interaction is among vehicles belonging to the same flow. It is the main type of flows in freeways and facilities where there are no intersections or other external factors.

- *Interrupted flows*: Traffic signal controls, stop or yield signs and other types of control devices influence the progression of vehicles along the section. Flows driving along these sections are said to be interrupted. Congestion may occur because of frictions internal to the traffic stream, because of interactions among streams and also because of the way traffic is regulated.

The driving time represents the main determinant in all uninterrupted flow sections. This characteristic is determined in time by tracing the position of each vehicle in between two sections. The difficulty to derive the instantaneous speed for each vehicle motivates the use of traffic flow models based on average conditions (macroscopic approach). In these models traffic states are represented in a road section by means of descriptive parameters like average flows, densities and speeds. Travel times are then estimated without consideration of each single vehicle trajectory but they are representative of all vehicles in the section. After Greenshields' fundamental diagram (Greenshields, 1935), which assumed a simple linear relationship between speed and density, several analytic relationships between speeds, densities and flows have been proposed to model traffic with more realistic assumptions (e.g. (Lighthill, 1955), (Greenberg, 1959), (Drew, 1965), (Van Aerde, 1995)).

The need for travel time models that directly relate the state variables (i.e. flows, saturation flows, etc.) in a simple formulation have motivated the development of heuristic time dependent models (e.g. (Davidson, 1966) or (Akcelik, 1991)), which are particularly appealing in design and planning problems. At interrupted flow sections the speed of vehicles has little role since the travel time is primarily influenced by the waiting time at the junctions. The fundamental role of traffic delays and queues is confirmed by its use in the computation of the level of service (e.g. in the Highway Capacity Manual (TRB, 2000), in the design of infrastructures, and in the estimation of environmental costs (e.g. fuel consumption and gas emissions). Performance assessment is based on assumptions regarding the characterization of the traffic arrival and service processes (Rouphail, 2000).

Interrupted flow sections are subdivided into non-signalized and signalized intersections, according to the type of regulation adopted. The service time on the first type depends on the probability for a driver to have enough gap between vehicles of the conflicting streams to pass the intersection safely. The service time of second type is instead determined by the control system, which controls the section. The remainder of this chapter deals with the development of delay and queuing models at signalized intersections, which give foundations to the development of the methods presented later in this thesis. For an extensive overview of queue and delay models at non-signalized intersections one can refer for example to Troutbeck and Brilon (Troutbeck, 2000).

3.2.1 Overview of modeling approaches for signalized intersections

The most important traffic performance, which determines the functionality of signalized intersections, is the *signal delay*. This characteristic is usually defined as the extra (waiting) time a traveler experiences due to the signal. This measure is therefore defined as the difference between the time spent to be served at the signal and the time the driver would have spent if the road section was uninterrupted. Section 3.3 deals with the delay models which have been developed in the last 50 years. Focus is given to the analytic models while Section 3.7 discusses the opportunity to use simulation. Since signal delay is partly caused by queues forming upstream the signal, the modeling of these queues represents also an important area of research, which is described in Section 3.4. Particular interest is given to the modeling of *overflow queues*, which occur when the number of arrivals is larger than the departures during one cycle, or during more cycles. A third important characteristic at signalized intersections is the signal capacity, which represents the (possible or probable) number of vehicles that can be served within an interval of time, e.g. a cycle or an hour. This characteristic is usually input for analytic delay and queue models and for macroscopic simulation, while it is an output measure in microscopic simulation programs. Since this thesis deals primarily with analytic and macroscopic simulation based models the modeling of this characteristic is not covered in this state-of-the-art. Critiques to the queue and delay models described in this chapter have been discussed in Section 3.5.

The modeling of delays and queues depends also on the type of application it is developed for. Therefore, different models are developed for *isolated intersections*, i.e. intersections whose performance does not depend on the performance of other signals upstream. Classification should be done also for models for fixed controls and dynamic controls. The effect of upstream signals is further discussed in Section 3.6. For sake of clarity, it is important to differentiate the models described in this chapter also from three other perspectives:

- *Assumptions on the process*: a model is usually *static* or *dynamic* depending on the assumption of stationariety or non-stationariety for the arrivals and/or the departure process.
- *Level of detail*: depending on the level of detail of the input and the output characteristics a model can be classified among *macroscopic*, *mesoscopic*, and *microscopic*. The first type deals with input and output characteristics as aggregated, i.e. link or route flows. The second type usually treats macroscopic characteristics as probabilistic variables. Finally the third level deals with the modeling of the movement of individual vehicles interacting one with another.
- *Modeling technique*: models can be finally classified into *stochastic* (or *probabilistic*) or *deterministic* whether they take or not into account the variability of the input and the output variables. Models can be further classified into analytic or simulation based if they are described or not by direct relationships between the input and the output variables.

This systematic classification inspires the following Table 3.1, which frames the different models presented throughout this chapter.

Table 3.1: Overview of models presented in this chapter

Level of detail	Model	Appl. type	Ass. process	Output type	Model tech.
	MACROSCOPIC				
	Beckman et al. (1956)	if	s	W	AP
	McNeill (1968)	if	s	W	AP
	Webster (1968)	if	s	W	AD
	Clayton (1941)	if	s	W	AP
	Miller (1963) – (1968)	if	s	Q	AP
	Newell (1965)	if	s	QW	AP
	May and Keller (1967)	if	d	W	AP
	Kimber and Hollis (1979)	if	d	W	AD
	Miller (1968)	if	s	Q	AP
	Akcelik (1980) – (1993)	if	sd	QW	AP
	Rouphail (1992)	af	s	W	AP
	Van As (1990)	af	s	QW	AP
	Tarko and Rouphail (1990)	af	d	Q	SD
	Stephanopoulos et al (1979)	if	s	W	AD
	Lin and Mazdeysa (1983)	it	s	W	AD
	HCM (TRB, 2000)	iaft	d	W	AP
	CTM (Daganzo, 1994)	aft	d	Q	SD
	LTM (Yperman, 2006)	aft	d	W	SD
	Bliemer (2007)	a	d	Q	AD
	MESOSCOPIC				
	Van Zuylen (1985)	if	d	Q	SP
	Olszewski (1990) – (1994)	if	d	QW	SP
	Brilon and Wu (1990)	if	d	W	SP
	MICROSCOPIC				
	INTEGRATION (Van Aerde, 1995)	ifat	d	QW	SP
	VISSIM (PTV, 2003)	ifat	d	QW	SP
	AIMSUN (Barcelo, 2003)	ifat	d	QW	SP

i=isolated a=arterial corridor f=fixed t=time-dependent control s=static d=dynamic W=delay Q=queue A=analytic S=simulation D=deterministic P=probabilistic

3.3 Delay models at signalized intersections

Traffic control on intersections is done by means of traffic lights that are most of the time operated automatically using a cyclic sequence of green, amber and red lights. The timing of the green and red duration can be fixed and predetermined (*fixed time* or *pre-timed* control). Another alternative is *semi-actuated* control. The main flow is interrupted after a green time of at least the minimum green time if the detector on the side road is activated. The green time on the minor road is determined by gap measurements, i.e. the green phase terminates when the gap between two vehicles is larger that a certain maximum. *Fully actuated control* is the mode of operation where all approaches have detectors and all green phases are controlled by means of detector information. *Demand responsive* control is a method in which green phases are only shown if there is demand and the length is also determined by demand as in fully actuated control.

The choice of the control type and the determination of the optimal control phases to adopt at one intersection are mainly done with the objective of reducing the delay to the vehicles. Delay is usually defined as the difference between the travel time experienced by a vehicle passing an intersection and the travel time experienced if the vehicle passes the intersection at cruise speed. With this definition delays are determined by waiting times because of signal operations and queues, and by lost times due to individual vehicle acceleration and deceleration characteristics. Delays are commonly referred to as *control delays* if one considers the delay due to only signal operations and queues, and *stopped delays* if also lost time due to acceleration and deceleration is computed.

Measurements of delays, stops, queue length etc. can be done using video, automatic detectors or visual observations. Such measurements give a more or less approximate estimate of the actual (or past) traffic conditions (*data-driven approach*). Alternatively, one can use the models presented in this chapter, which allow a prediction of the future conditions of traffic, or the estimation of expected effects of changes in the network status (*model-based approach*). The state-of-art presented in the following of the chapter deals primarily with the component of delays caused by signal operations and queues, while little importance is given to the acceleration and deceleration effects. As a consequence, the following state of the art deals exclusively with *control delay models*.

3.3.1 Basic concepts

Vehicles arriving during the green phase pass the stop line without delay if there is no queue to be served waiting at the intersection. Vehicles arriving during the red phase, and while a queue needs still to be dissolved, experience a delay, which depends on the signal phase plan, on the capacity of the road sections and on the traffic flow conditions. During the red phase, the arriving vehicles queue up, while they will be served within the next or later green phases. Figure 3.1 illustrates schematically an example of queuing process and its relationship with the delay within a cycle and for one traffic stream.

The three signal phases (green, amber and red light) are, in the scheme, simplified into two phases, effective green time t_g and effective red t_r. The effective green time is the green time from which the *green start lag* is subtracted and a *green end lag* is added. The *green start lag* is partly due to the reaction time of the first driver passing the intersection during green, but most of the time is the consequence of the vehicles' acceleration operations, which make the speed of the first few vehicles passing the signal lower than in the middle of the green phase. At the end of the green phase, during the amber, vehicles still enter the intersection. The average time that the amber phase is still used by vehicles entering the intersection is called the *green end lag*. Using the concept of effective green time, the flow is assumed to reach instantaneously the maximal saturation flow rate s at the moment the signal turns green, and to stop when it turns red. Therefore, if one assumes constant arrival rate q and constant departure rate s the queue will increase linearly during the red phase and, if $q < s$, it will decrease till the end of the effective green phase, otherwise it will still increase but with a rate of $q - s$.

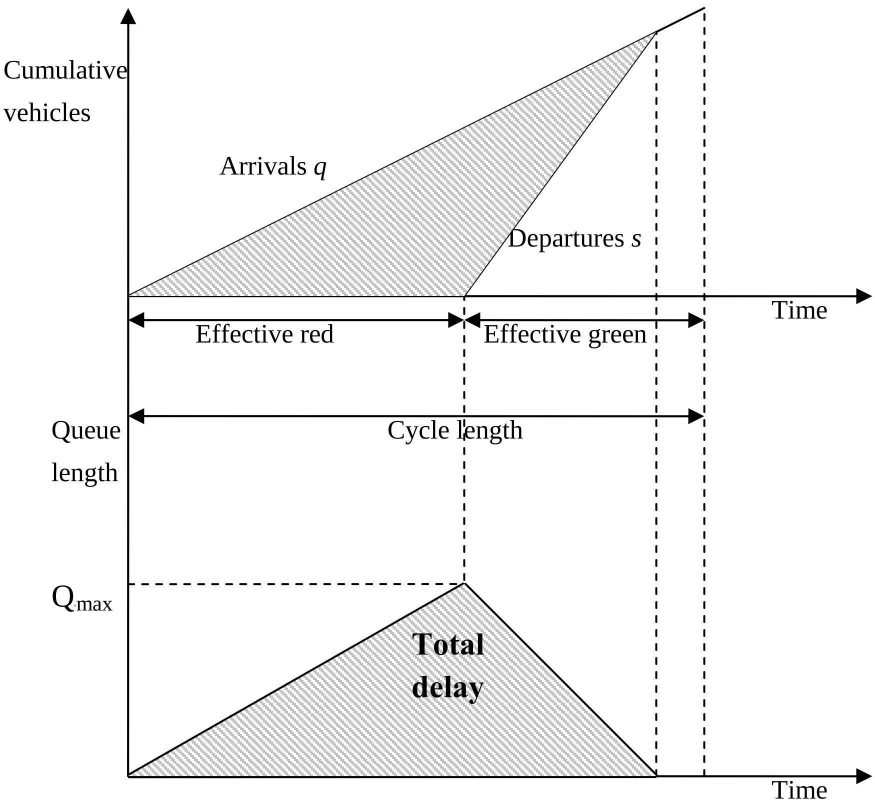

Figure 3.1: Relationship between arrivals, departures, queue length and total delay in undersaturated intersections with uniform arrivals and departures

In reality vehicles do not generally arrive at constant rates, and also the number of departures may be different from cycle to cycle due to vehicle composition, drivers' reaction skills etc. If a signal is undersaturated, as in Figure 3.1, the arrivals during the red phase and accumulating at the stop signal are likely to be served within the following green phase.

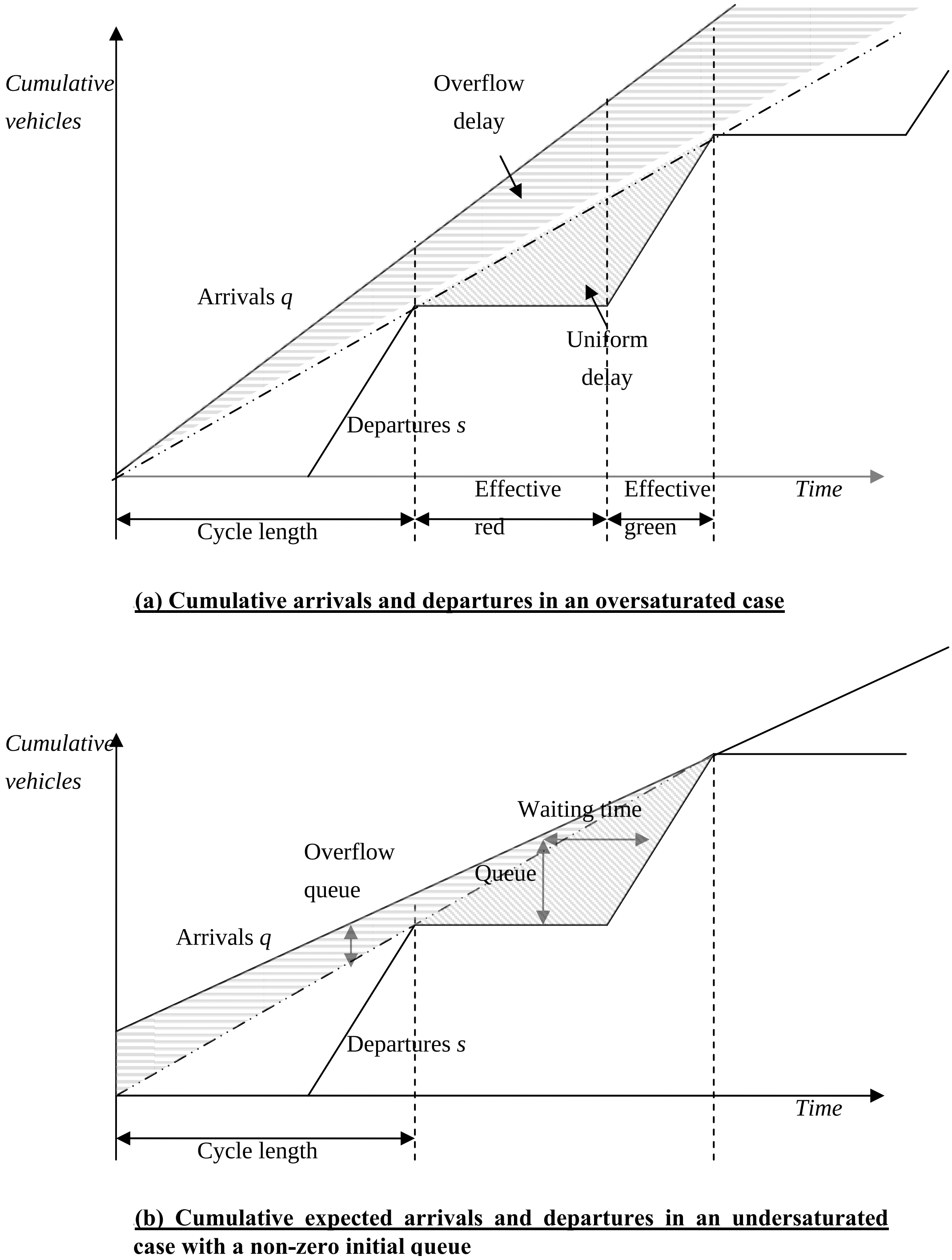

Figure 3.2: evolution of arrivals and departures with overflow delay

If arrivals and departures are assumed arriving uniformly in time the delay is simply computed by the area of the triangle bordered in bold in Figure 3.1. This area computes the so-called *uniform delay* component in an undersaturated signal. If an intersection is oversaturated, the assigned green phase is not sufficient to serve all vehicles arriving during the cycle and a residual queue will occur. A residual queue may be observed also if an undersaturated period follows an oversaturated one and a non-zero initial value is assumed. The residual queue is usually called *overflow queue* and the corresponding delay *overflow delay*. Figure 3.2 schematizes the queue evolution in the assumption of uniform arrivals and departures within a cycle for both cases, showing the evolution of the overflow queues and delays.

Figure 3.2 (a) illustrates the oversaturated queue with the assumption of zero initial queue and average flow larger than the capacity. The shaded areas show uniform delay and overflow delay generated by the overflow queue. The same applies to Figure 3.2 (b), which shows the undersaturated case with a positive initial queue length. While in the first graph the overflow queue increases with cycles, in the second case the initial queue will generate extra delay to the vehicles arriving at following cycles until the line representing the cumulative arrivals intersects the one of the cumulative departures. After this moment the intersection behaves like in Figure 3.1 and the only component of delay is represented by the uniform one.

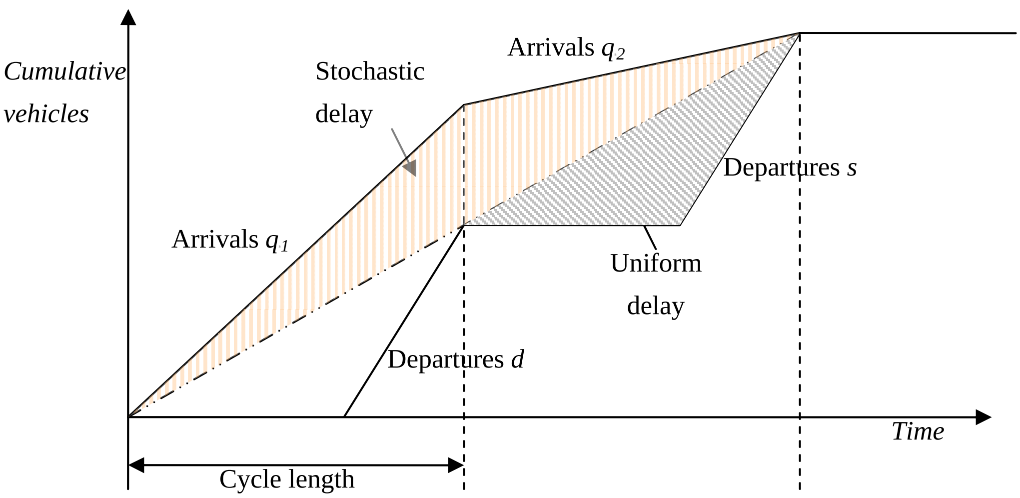

Figure 3.3: overflow delay caused by stochastic arrival process

Chapter 2 has shown that vehicle headways are unlikely to be uniformly distributed in time but stochastic fluctuations can be observed. If a period of stationary arrival rate is assumed, the arrivals at each cycle can vary around this value. Therefore, it might be likely that one or more cycles are unable to serve all arrivals within the assigned green times but one or more cycles may also be needed to serve part of these vehicles. In this case an extra delay is computed together with the uniform one (*stochastic delay*). Figure 3.3 gives an example of this temporary queue and the consequent stochastic delay.

Stochastic queuing can also occur because of the stochastic behavior of the service rate. Departures can vary in time due to several reasons, e.g. variable reaction times, different car-following criteria among drivers etc. The smaller the flow to capacity ratios the more likely that each cycle starts and ends with a zero overflow queue and the stochastic character of the demand plays little role in the estimation of delays. In these conditions the queue length can be considered stationary over cycles and steady-state delay models can be applied with little error. As traffic intensity increases, there is an increasing likelihood that some cycles will begin or end with an overflow queue of vehicles even when the average flow is smaller than the capacity. The stochastic component in this case can also be considerably larger than the uniform component and represents the main cause of delays. The closer the demand to the capacity, the longer time is needed to dissipate the effect of these queues.

At extremely congested conditions, the stochastic queuing effects are minimal in comparison with the size of oversaturated queues and deterministic models based on fluid theory have been demonstrated to be appropriate. Undersaturated periods following oversaturated ones are not yet sufficiently investigated if the demand is close to the signal capacity. Considering that most real-world signals are desired to operate as close as possible to the capacity, time-dependent models are of particular relevance for this range of conditions (Rouphail, 2000). For this reason several approximate expressions can be found in literature for both steady-state and dynamic models, which form the basis for the formulas suggested by most of the capacity guides. Some examples of these expressions are given in this chapter.

3.3.2 Steady-state delay models

Steady-state delay models are developed under the assumption of stationary conditions for the overflow queue length, thus they are applicable only to undersaturated cases. Exact expressions can be derived from mathematical relationships under some simplifying assumptions, as described in subsection 3.3.2.1. The gap between exact expressions and field data justifies the development and adoption in practical studies of several approximate expressions, as described in subsection 3.3.2.2.

3.3.2.1 Derivation of exact expressions

The seminal work of Beckmann et al. (Beckmann, 1956) includes the first derivation of expected delay at isolated traffic signals with the assumption of constant service rate and Binomial arrival distribution. The expression derived is the following:

$$E[W] = \frac{t_r}{t_C \cdot (1 - q/s)} \cdot \left[\frac{E[Q_O]}{q} + \frac{t_r + 1}{2} \right] \tag{3.1}$$

With:

- $E[W]$ expectation value of the delay (sec),
- t_C signal cycle,
- t_r length of the effective red,
- q arrival flow rate,
- s saturation flow and
- $E[Q_O]$ expectation value of the overflow queue length (in vehicles).

The expectation value of the delay is thus in linear relationship with the expectation value of the overflow queue length $E\{Q_O\}$, which should also be estimated.

McNeil (McNeill, 1968) derived an exact formula from general arrival distributions by assuming the delay as sum of two components: $W = W_1 + W_2$ where W_1 is the total vehicle delay experienced during the red phase and W_2 during the green phase. The components (expressed in veh*sec) are computed with the following integrals:

$$\begin{aligned} W_1 &= \int_{t=0}^{t=t_r} \{Q(0) + A(t)\} dt \\ W_2 &= \int_{t=t_r}^{t=t_C} Q(t) dt \end{aligned} \tag{3.2}$$

Where $Q(0)$ is the assumed initial queue length at the starting of the red phase and $A(t)$ is the cumulative arrival distribution within the red phase. During the green phase the delay is only a function of the queue length in time $Q(t)$. Taking the expectations of Formulas (3.2) one can derive the expectation value of delay. Assuming the arrivals as stationary (with average rate q) the first component can be expressed as:

$$E[W_1] = t_r \cdot E[Q(0)] + \frac{1}{2} q \cdot t_r^{\,2} \tag{3.3}$$

The total vehicle delay experienced during the green phase W_2 is derived by computing the total waiting time for a queue $Q(t)$ to be fully served. Therefore, if one defines Z_2 as a random variable expressing the total delay experienced during green when the cycle t_C

is infinite, the total waiting time in a busy period for a queuing process with arrival rate q and deterministic service time $1/s$ and initial system state $Q(0)$ is expressed by:

$$E[Z_2] = \frac{(1 + I \cdot q/s - q/s) \cdot E[Q(0)]}{2 \cdot s \cdot (1 - q/s)^2} + \frac{E[Q^2(0)]}{2 \cdot s \cdot (1 - q/s)} \quad (3.4)$$

In formula (3.4) the variability of flows is considered introducing the following index of dispersion:

$$I = \sigma(A) / q \cdot t_C \quad (3.5)$$

The second component of the delay function can thus be then expressed by:

$$\begin{aligned} E[W_2] &= E[Z_2 \mid Q(t = t_r)] - E[Z_2 \mid Q(t = t_C)] \\ &= \frac{(1 + I \cdot q/s - q/s) \cdot E[Q(t = t_r) - Q(t = t_C)]}{2 \cdot s \cdot (1 - q/s)^2} + \frac{E[Q^2(t = t_r) - Q^2(t = t_C)]}{2 \cdot s \cdot (1 - q/s)} \end{aligned} \quad (3.6)$$

So far, no assumptions have been made on the dynamics of the queue length. The above expression is thus valid also when the queue length is not in stationary conditions. If instead one assumes stationary queue length as much as arrivals, skipping some arithmetic manipulations (see (Rouphail, 2000) for more details) one arrives at the exact expression of the average vehicle delay:

$$E[W] = \frac{t_r}{2 \cdot t_C \cdot (1 - q/s)} \left\{ t_r + \frac{2}{q} E\{Q(0)\} + \frac{1}{s}\left(1 + \frac{I}{1 - q/s}\right) \right\} \quad (3.7)$$

This model is valid *only* if the mean of the queue length is stationary in time and under demand in steady-state conditions. For non-stationary queue conditions, the expected delay can only be computed using Formula (3.6), which still requires the knowledge of the expectation value of the queue at the start and end of the green phase. This justifies why the research has moved towards the research of exact expressions for the expected value of the queue length instead of directly calculating delays.

3.3.2.2 Approximate expressions

Webster (Webster, 1958) was the first to propose an approximate expression for the delay formula combining formulas from theoretical relationships with results of simulations:

$$W = \frac{t_C \cdot (1 - t_g / t_C)^2}{2 \cdot [1 - (t_g / t_C \cdot) \cdot x]} + \frac{x^2}{2 \cdot q \cdot (1 - x)} - 0.65 \cdot \left(\frac{t_C}{q^2} \right)^{\frac{1}{3}} \cdot x^{2 + 5(t_g / t_C)} \tag{3.8}$$

where $x = q / c$ is the degree of saturation and $c = s \cdot t_g / t_C$ is the signal capacity. The first term is the analytical derivation of uniform delay, while the second is a characterization of random or stochastic delay, analytically derived assuming Poisson arrivals and deterministic service rate. The last term has been introduced to reduce the discrepancy with results observed from simulation data and it is thus purely empirical. This term is frequently assumed to be around 10% of the first two terms, thus Formula (3.8) is often simplified with the following expression:

$$W = 0.9 \left\{ \frac{t_C \cdot (1 - t_g / t_C)^2}{2 \cdot [1 - (t_g / t_C) \cdot x]} + \frac{x^2}{2 \cdot q \cdot (1 - x)} \right\} \tag{3.9}$$

This formula is still widely used in practice and it is frequently used as benchmark for the derivation of approximate expressions assuming different arrival processes and optimal signal settings. An approximation of the exact Formula (3.7) has been proposed even earlier by Clayton (Clayton, 1941) in the assumptions of LIFO discipline (Last In, First Out) and sufficiently heavy traffic conditions:

$$E[W] = \frac{t_C \cdot (1 - t_g / t_C)^2}{2 \cdot (1 - q / s)} + \frac{E[Q(0)]}{q} \tag{3.10}$$

As said, the exact expression derived by McNeill (McNeill, 1968) reprinted in Formula (3.7) is subject to the derivation of a sufficiently accurate estimation of the expectation value of the overflow queue. Initially research was concerned with deriving upper bounds of overflow queues instead of expected values. Miller (Miller, 1963) proposed to use the following Formula (3.11) which computes an upper bound for the value of $E[Q(0)]$ starting from the analytical true relationship:

$$E[Q(0)] = \frac{\sigma^2[c - q] - \sigma^2[\Delta c]}{2 \cdot E[c - q]} \tag{3.11}$$

where Δc is the reserve capacity in one cycle: $c - Q(0) - q$ if $Q(0) + q < c$, 0 otherwise.

Since Δc decreases when $E[c - q]$ approaches zero, an upper bound is obtained by neglecting the second term of the numerator. Using the index of dispersion (3.5) one can obtain the following expression:

$$E[Q(0)] \leq \frac{I_a \cdot x}{2 \cdot (1-x)} \tag{3.12}$$

Newell (Newell, 1965) proposed another approximation of delay Formula (3.7), which includes a component related to the variability of arrivals I_a:

$$E[W] = \frac{t_C \cdot (1-t_g/t_C)^2}{2 \cdot (1-q/s)} + \frac{E[Q(0)]}{q} + \frac{(1-t_g/t_C) \cdot I_a}{2 \cdot s \cdot (1-q/s)^2} \tag{3.13}$$

Moreover, Newell proposed in the same work a modification of Miller's queue Formula (3.12) in order to obtain the expectation of the overflow queue instead of the upper bound by simply adding a multiplicative factor:

$$E[Q_O] = H(\mu) \cdot \frac{I_a \cdot x}{2 \cdot (1-x)} \tag{3.14}$$

where

$$\mu = \frac{s \cdot t_g - q \cdot t_C}{\sqrt{I \cdot s \cdot t_g}} \tag{3.15}$$

The multiplicative factor $H(\mu)$ was provided by Newell in a graphical form, while Cronje (Cronje, 1983) gave an empirical approximation using simulation data:

$$H(\mu) = \exp\left[-\mu - \left(\frac{\mu^2}{2}\right)\right] \tag{3.16}$$

More examples of overflow queue models under steady-state conditions are given later in Section 3.4.

3.3.3 Time-dependent delay models

The methods described in Section 3.3.2 can give approximate estimates of steady-state delays and queues in one cycle but no information on the duration of such states, and consequently on the dynamics of delays can be made, with those approaches. An infinite period of stable traffic conditions is required, which is clearly an unrealistic assumption. Equilibrium models can be therefore acceptable only for low flow to capacity ratios; the closer the flow to the capacity, the longer time should be needed to reach equilibrium.

However, it is unlikely that above certain demand conditions one can observe evaluation periods, which are long enough to observe equilibrium queues. Alternative approach is to assume stationary demand and capacity for the entire evaluation period, while the resultant queuing process is non-stationary.

Alternatively, the time-dependent arrival profile of the overflow queue length can be subdivided into fractions of the whole evaluation period where it is assumed stationary (using e.g. step-wise demand, parabolic or triangular shapes) and compute the queue as stepwise too. If an expression of arrivals and departures in time is provided one can adopt the method given by May and Keller (May, 1967) for the computation of queues in time:

$$
\begin{aligned}
&E[Q(t)] = E[Q(0)] + A(t) - D(t) \\
&A(t) = \int_{\tau=0}^{t} q(\tau)d\tau \\
&D(t) = \int_{\tau=0}^{t} c(\tau)d\tau
\end{aligned}
\tag{3.17}
$$

An estimation of average vehicle delay within a period of length T is then given by:

$$
E[W(T)] = \frac{1}{A(T)} \int_{t=0}^{T} Q(t)dt \tag{3.18}
$$

The authors showed the results of this method by considering a triangular shaped arrival profile and constant capacity. However, results using this approach do not explain the delay propagation in conditions near capacity because of the strong influence of arrival and departure variability.

Heuristic methods were proposed in order to fill the gap in delay models when demand is near to the capacity. Kimber and Hollis (Kimber, 1979) used the coordinate transformation technique to adapt static models to give them some way of a time dependency. The coordinate transformation technique was addressed as a possible method for the derivation of time-dependent models of delay starting from the assumption that total delay is nearly equal to the uniform component W_1 (the first term of Equations (3.8), (3.9), (3.10) and (3.13)) for very low degrees of saturation while it is asymptotically equal to the following deterministic expression in cases of highly congested situations:

$$
W(T) = W_1 + \frac{T}{2} \cdot (x - 1) \tag{3.19}
$$

To model the transition between uniform delay experienced at undersaturated conditions and Formula (3.19) at oversaturated conditions, Kimber and Hollis applied the coordinate transformation technique to the queuing process. The theoretical evolution of expected value of the queue length under steady-state conditions assumes that queues tend asymptotically to infinity when the demand approaches capacity. The behavior of queues in oversaturated conditions follows instead an evolution that follows Formulas (3.17). If arrivals and departures have a constant rate and they are uniformly distributed, Formula (3.17) can be rewritten as:

$$E[Q(t)] = E[Q(0)] + (x-1) \cdot c \cdot t \tag{3.20}$$

The deterministic evolution of queues in undersaturated conditions is thus linear if demand and capacity are stationary. Kimber and Hollis assumed that, in conditions near the capacity and with a finite evaluation period T, the evolution of queues is a combination of the two trends, as shown in Figure 3.4. To combine the two different behaviors the authors simply mathematically transformed the vertical asymptote of the static formulas with the linear function (3.20).

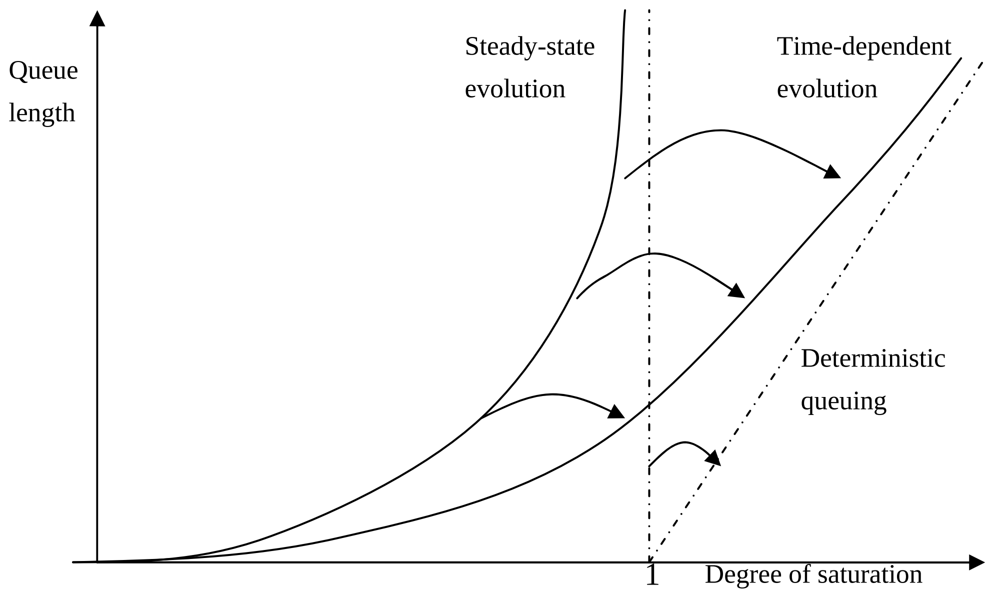

Figure 3.4: Coordinate transformation of the overflow queue model

The heuristic rule applied in this transformation is to rotate clockwise the vertical asymptote of the static queuing process until it coincides with the linear deterministic queuing model. By doing this, both queue and delay models have been approximated by a formulation in the form $1/2 \cdot [(a^2 + b)^{1/2} - a]$. The expressions of a and b have been provided by the same authors:

$$a = (1 - x) \cdot t_c \cdot T + 1 - E[Q(0)]$$
$$b = 4 \cdot \left\{ E[Q(0)] + x \cdot t_c \cdot T \right\} \quad (3.21)$$

Although this approach overcomes the gap between static steady-state models and deterministic time-dependent models, doubts still surround this approach since the coordinate transformation technique does not represent a theoretically valid approach but only a heuristic.

3.3.4 Capacity guides

The research in traffic flow modeling at signalized intersections is still developing a general theory, which encloses and describes the queuing and delay processes with realistic traffic states. Exact models sometimes require very elaborate computations, as for the case of the delay model (3.6). On a parallel track, transportation practice trades off model exactness with model handiness allowing the approximation that characterizes any model presented so far. Several practical models have been collected in the official transportation manuals. The most frequently used manuals are the American Highway Capacity Manual (TRB, 2000), the Canadian Capacity Guide (ITE, 1995), the Australian Capacity Guide (ARR, 1995).

3.3.4.1 Highway Capacity Manual 2000

The Highway Capacity Manual 2000 measures the performance of a signalized intersection by computing the expected stopped delay per vehicle. According to the Manual this delay d can be decomposed into three terms:

$$W = W_1 \cdot PF + W_2 + W_3 \quad (3.22)$$

where

- W_1 is the uniform stopped delay per vehicle (s/veh)
- W_2 is the incremental, or random stopped delay (s/veh)
- W_3 is the initial queue delay
- PF is the progression factor, to account for signal coordination

The first two delay components are given by the following formulas:

$$W_1 = 0.5 \cdot t_C \cdot \frac{\left(1 - t_g / t_C\right)^2}{\left(1 - \min(1, x) \cdot t_g / t_C\right)} \quad (3.23)$$

$$W_2 = 900 \cdot T \cdot \left[(x-1) + \sqrt{(x-1)^2 + \frac{8 \cdot k \cdot I_f \cdot x}{c \cdot T}} \right] \qquad (3.24)$$

where the parameter k is respectively 1 for fixed time controls, 0.5 for semi-actuated signals and in between 0.04 and 0.5 for actuated controls). I_f is the filtering adjustment factor. The stochastic component is similar to the delay formulation provided by Kimber and Hollis. Accordingly, this formula assumes the queue length to be constant and finite if $x < 1$ while it behaves according to the linear deterministic function for $x > 1$.

The third component is computed by specifying the parameters of the formula:

$$W_3 = \frac{1800 \cdot Q(0) \cdot (1+u) \cdot t}{c \cdot T} \qquad (3.25)$$

A procedure to derive the values of the parameters u and t can be found in the Appendix F of chapter 16 of the Manual. The initial queue $Q(0)$ is computed using the linear deterministic model (3.20), therefore it considers a positive value only if the previous period was oversaturated.

3.3.4.2 The Canadian Capacity Guide

The Canadian Capacity Guide considers only uniform and incremental delay components, while no initial queue delay is present:

$$W = W_1 \cdot k_f + W_2 \qquad (3.26)$$

with k_f replacing PF of Formula (3.22) but very little difference can be found comparing the computation of the two parameters. While the expression of W_1 coincides with Formula (3.23), the incremental delay is given by:

$$W_2 = 15 \cdot T \cdot \left[(x-1) + \sqrt{(x-1)^2 + \frac{240 \cdot x}{c \cdot T}} \right] \qquad (3.27)$$

3.3.4.3 The Australian Capacity Guide

Similarly to the Canadian Guide, the Australian Capacity Guide does not consider a delay due to a positive initial queue and has the same Expression (3.23) for the uniform delay component, but it does not consider the progression factor:

$$W = W_1 + W_2 \qquad (3.28)$$

The incremental delay has another analogous formulation to Formula (3.24):

$$W_2 = 900 \cdot T \cdot \left[(x-1) + \sqrt{(x-1)^2 + \frac{m \cdot (x - x_0)}{c \cdot T}} \right] \tag{3.29}$$

In case of uniform arrivals $m = 12$ and the analogy becomes also evident with the Akcelik's queue formula presented in the next section.

The newly introduced issue of estimating the effect of oversaturated periods on the following periods, proved by the recent introduction of the initial queue delay in the HCM only in its latest version, can indicate that research on this direction is a key issue and motivates the research throughout this thesis.

3.4 Analytic queue models at isolated fixed time signals

Queuing models are needed to evaluate the delay the travelers experience, for example using Formula (3.7), but it can also be helpful to evaluate other characteristics, which cannot be easily assessed evaluating expected delays. For example, spillback effects and the evaluation of the length of exclusive turning lanes cannot be modeled without the estimation of the queue length and its variability.

Concerning the steady-state expression several approximations of Formula (3.14) were proposed in the last decades for the computation of the exact delay Formula (3.7). Miller (Miller, 1968) for example proposed the following simplification:

$$Q_O = \frac{\exp\left[-1.33 \cdot \sqrt{s \cdot t_g \cdot \frac{1-x}{x}} \right]}{2 \cdot (1-x)} \tag{3.30}$$

Akcelik (Akcelik, 1980) further simplified this expression with the following:

$$Q_O = \begin{cases} \frac{1.5 \cdot (x - x_0)}{1-x} & \text{when } x > x_0 \\ 0 & \text{otherwise} \end{cases} \tag{3.31}$$

The parameter x_0 represents the value above which overflow queues can be considered non-negligible and it has the following expression:

$$x_0 = 0.67 + \frac{s \cdot t_g}{600} \tag{3.32}$$

These formulas assume an infinite time period for stable traffic conditions to be achieved. In these ideal conditions, if the degree of saturation is equal or larger than one, the expected equilibrium value of the overflow queue is infinite. If the degree of saturation is low, the equilibrium value of the overflow queue is reached after a few cycles and steady-state formulae can still be a valid approximation. If the degree of saturation is nearly one it may take too long to reach the equilibrium and it is more opportune knowing the behavior of the queues at the end of a pre-determined period.

Kimber and Hollis' approach represents a good approach in this sense since it computes both characteristics and gives them a dynamic feature. Their model is although derived under the assumption of zero initial queues, constant degree of saturation within a fixed time period. Despite these limiting assumptions, this approach is still very appealing, and for this reason their model is still widely applied. For example this model is used in the TRANSYT program (Robertson, 1980).

Akcelik (Akcelik, 1980) formulated the expression that is most frequently used by practitioners for the expectation value of the overflow queue in time. He provided a formulation of the queue evolution using also the coordinate transformation technique:

$$Q_O = \begin{cases} Q_O(T) = \dfrac{c \cdot T}{4}\left(x - 1 + \sqrt{(x-1)^2 + \dfrac{12 \cdot (x - x_0)}{c \cdot T}} \right) & \text{when } x > x_0 \\ 0 & \text{otherwise} \end{cases} \tag{3.33}$$

Similarly to the Kimber and Hollis model, this model has the property to give time-dependency to the expectation of the overflow queue and to estimate such queues in a fixed time period T in both undersaturated and oversaturated conditions. Regarding the computation of delays, Akcelik proposed another expression for Formula (3.14):

$$W = \begin{cases} \dfrac{t_C \cdot (1 - t_g / t_C)^2}{2 \cdot (1 - q / s)} + \dfrac{Q_O}{c} & \text{when } x < 1 \\ \dfrac{(t_C - t_g)}{2} + \dfrac{Q_O}{c} & \text{when } x \geq 1 \end{cases} \tag{3.34}$$

This formula is probably the most used model in case of flows approaching capacity and initial queue $Q(0) = 0$. A recent extension of the former model for including the effect of an initial queue can be found in the aaSIDRA manual (Akcelik, 2002). The author adds a term to the flow rate q, which considers the residual queue generated in the previous intervals.

The model is assumed valid in ideal conditions of constant arrival and departure rates and empty signal at the start of the evaluation period. Moreover, the model has been shown to well represent overflow queues and delays when the evaluation period is fixed to 15 minutes, which is a typical interval in practical studies. On the other hand, Brilon and Wu (Brilon, 1990) showed, using a Markov Chain approach to generate the overflow queues, that the Akcelik model does not properly approximate the expected value of queues and delays if arrival rates are not constant. Starting from a parabolic shape of the demand in time, the authors develop a heuristic formula for the expected overflow queue as alternative to Akcelik's model.

3.5 Critiques to analytic queue and delay models

The many approaches developed to model the behavior of traffic at signalized intersections, and presented in this chapter, reflect the lack of a general theory, which encompasses the various dynamic and stochastic aspects of queues and delays under different conditions of traffic.

Firstly, main critique is in the way the previous approaches deal with the transition between uncongested and congested states. The gap between steady-state models and time-dependent models still remains a weak point. This issue affects the understanding of the dynamics of traffic in e.g. peak period analysis. A methodology that is able to catch the smooth transition between the dynamics of traffic in undersaturated conditions and oversaturated ones will help in filling this gap.

3.5.1 Dynamic and stochastic behavior

Large inconsistencies have been found between the delay models of Webster (Webster, 1958), McNeill (McNeill, 1968) and Miller (Miller, 1968) among others and with field data (Ohno, 1978). The main reasons for these inconsistencies stand in the dynamic and stochastic behavior of traffic, and the way the flows propagate along the network, which rise serious doubts on the assumption of stationary conditions assumed by all steady-state queue models in undersaturated conditions. The use of the coordinate transformation technique is a convenient, but not rigorous way of solving the problem. For this reason several modifications of the above queue and delay models have been proposed so far by simply adding multiplicative factors to account for the variable profile of traffic (e.g. (Rouphail, 1992), (Akcelik, 1993), (Fambro, 1993)).

It appears that the dynamics of the overflow queue is strongly affected by the dynamic and stochastic properties of the arrivals and the departures. Especially in conditions of demand near capacity the overflow queues show a strong stochastic behavior and the standard deviation can be of the same magnitude as the mean, making the prediction of expected travel times very difficult (Van Zuylen, 2003). The strong stochastic behavior of queues in conditions near capacity justified a stochastic modeling approach, the Markov Chain model, adopted in the past by several authors. Chapter 4 and Appendix A describe in detail this methodology which has inspired consistently the models developed in this thesis. Among others, Van Zuylen (Van Zuylen, 1985), Olszewski ((Olszewski, 1990a), (Olszewski, 1990b), (Olszewski, 1994)) and Brilon and Wu ((Brilon, 1990), (Wu, 1990)) used the Markov Chain technique to simulate the evolution of the overflow queue length at the end of a green phase in a stochastic modeling fashion. Apart from the work of Brilon and Wu, no other study attempts to derive an empirical formula for the expected value of the queue with non-stationary arrivals.

This methodology has been applied so far only in a cycle-to-cycle process, i.e. the queue length probability distribution at one cycle is determined by simply the overflow queue state at the previous cycle. This means that no information is given on the queuing process within the cycle. An implication of this lack is for example in the estimation of maximum queue lengths in a cycle. Gridlocks or spillback effects may be produced by these inner queuing process, e.g. in a multi-phased control system. Chapter 5 fills this gap by proposing a probabilistic model formulation also for the queuing process within each cycle.

3.5.2 Vertical vs. horizontal queues

The modeling of the expected value of the overflow queue using the analytic approach described in Section 3.4 assumes queues to build up 'vertically' (*vertical queuing analysis*), meaning that they do not deal with the physical space occupied by the vehicles but they are only interested in their expected number at one point in time. This assumption does not allow any investigation of spillback effects of this queue on other links or nodes. This assumption affects the computation of the link travel time functions, which are often considered fixed in e.g. dynamic network loading processes (Bliemer, 2007).

Higher level models like microscopic simulation programs enable the analysis of the queuing process in great detail allowing *horizontal queuing analysis*. Spillback can be accounted for by these models, but at the price of long computation times, yet unbearable for some transportation problems (e.g. traffic assignment, signal optimization etc.).

Some macroscopic simulation-based models can also deal with horizontal queues. Examples of this modeling approach applied to the queuing process are the Cell

Transmission Model (CTM, e.g. (Daganzo, 1994)) and the new Link Transmission Model (LTM, see (Yperman, 2006)), which account for the dynamics of queuing by subdividing a link into discrete cells.

The dynamic propagation of queues in time and space, and therefore the behavior of queues within each cycle, can also be computed using shockwave theory. Apart from using microscopic models, traffic flow dynamics can also be computed analytically by means of aggregated characteristics like flows, speeds and densities of each road section. Using fluid-dynamic relationships, the dynamic propagation of the traffic flow and especially the congestion effects can be tracked on a network. Lighthill and Whitham (Lighthill, 1955) and Richards (Richards, 1956) successfully implemented fluid-dynamic theories to demonstrate the existence of shockwaves propagating in time and space in highways. Rorbech (Rorbech, 1968) extended these models to signalized intersections. Stephanopoulos and Michalopoulos (Stephanopoulos, 1979) and Michalopoulos and Stephanopoulos (Michalopoulos, 1980) demonstrated the existence and investigated the behavior of shockwaves at traffic signal caused by the periodic signal operations. The benefit of applying the shockwave method in the opportunity to deal with horizontal queues was shown in Michalopoulos and Pisharody (Michalopoulos, 1981), where the authors developed a control algorithm, which minimizes the total delay of a network taking into account maximum queue lengths allowed for each road section as constraints.

3.5.3 One-lane vs. multiple lanes

Road sections with more than one lane and with overtaking possibilities have consistently different behavior than single lane or restricted overtaking sections. The impact of different driving behavior of vehicles is counterbalanced by the possibility to overtake and reduce the interaction among vehicles. Lane changing possibilities are also subject to limiting factors like gap acceptance of vehicles or platooning effects. Multilane traffic flow modeling is an issue, which represents a large area of research in traffic flow theory at freeways. For detailed analytic description of the traffic phenomena in multilane road sections using macroscopic relationships one can refer to e.g. (Hoogendoorn, 1996), (Helbing, 1997), or (Ngoduy, 2006).

Lane changing behavior of travelers can strongly influence the dynamic and stochastic propagation of queues and delays in multilane signalized intersections. Lanes can be differently preferred by the travelers because of various reasons (e.g. stay on the right lane to favor the overtaking of a faster vehicle, or choose a lane, which favors an operation downstream, like turning, or parking etc.). This unbalanced distribution of flows among lanes can result in a different formation of queues at the intersection. The variability of such queues is thus dependent on the variability of the relative density among lanes. Lane changing behavior may reduce in some way this variability. Travelers can decide to change their current lane with another one if they expect a reduction of their waiting time at the signal. This can strongly modify the distribution of vehicles among service points, and upstream among lanes. Knowing how vehicles can reduce this variability of queues can for example be important if toll plazas are planned in a road section and the modification of the infrastructure upstream should be designed in order to minimize the total expected delay.

Very little research has been done so far on the performance of signalized intersections with multiple lanes per stream, and it is common in practice to consider flow equally distributed when approaching the intersection.

There are a few principles for estimating lane flows, which implicitly account for the lane choice of travelers:

- *Equal degree of saturation*: used in e.g. SIDRA (Akcelik, 2002). Drivers tend to distribute according to an equal utilization of lane capacity or with maximum throughput. This lane choice behavior is therefore static and based on travelers' perception of lane capacity. In practice this makes difference when the capacity of one lane is affected by some penalty factors (e.g. it is a shared lane).

- *Equal flow ratio*: based on flow-to-saturation flow ratios. This criterion is used in the Highway Capacity Manual and in OSCADY (Burrow, 1987). It differs from the first criterion only because it does not consider difference in effective green times among lanes.

- *Equal average delay*: lane choice is based on a delay minimization criterion (Bonneson, 1988).

- *Minimum travel time*: drivers always change lane according to a minimum route travel time criterion.

- *Equal queue length principle*: adopted by some microscopic models (e.g. INSECT (Cotterill, 1984), AIMSUN (Barcelo, 2003), VISSIM (PTV, 2003) etc.), drivers are assumed to choose always the lane with the shortest queue.

It is difficult to judge the validity of one or another criterion. While the first two seem "too simplistic", as they treat lane choice as static and irrespective of the environment surrounding the intersection, the others suffer of a lack of good queuing and delay models, which are often assumed under steady state conditions. The travel time approach was supported by Fisk (Fisk, 1988), who stated that "a minimum travel time principle should be used for traffic assignment and intersection calculations to achieve consistency between these two levels of modeling from a behavioral viewpoint". Akcelik (Akcelik, 1989), on the other hand, replied that distinction should be made between strategic and tactical decisions. At route choice level drivers may pre-select routes, while at intersections they are more sensitive to direct measures (e.g. capacity). Akcelik finally stated that “any principle has a shortcoming: they are all steady-state approximations of a dynamic process. Lane choice of drivers may change with the congestion level, and it can vary between red and green interval”.

Very recently, Tian and Wu (Tian, 2006) proposed a capacity estimation method for intersection approaches with a short flare that accounts for the dependence of the capacity underutilization on the arrival rate and its variability. They considered the effects of lane blockage by considering the distribution of traffic among lanes as stochastic and later they evaluated the influence of right-turn lane length, of the proportion of right-turn vehicles and the length of the cycle time to the signal throughput. The proposed approach computes the capacity value of the full approach and of the straight-through lanes and flares separately, but it does not compute the delay and the overflow queue incurred by this capacity reduction. Furthermore, there are two shortcomings in this approach: 1) the model considers only one lane dedicated to the through traffic, so no lane changing behavior is assumed and 2) the model does not consider the dynamics of the traffic arrivals in between cycles and therefore a residual overflow queue from previous cycles. Nevertheless, the model enables one to compute a fundamental input for the estimation of queues and delays at undersaturated approaches. Moreover its use in the equal degree of saturation principle should improve the lane flow estimation of programs like SIDRA by giving a demand-dependent flow distribution.

It is still unclear how to model the selection criterion of travelers (or consumers in general) in a dynamic scenario. Jumsan et al. (Jumsan, 2005) recently analyzed the lane changing behavior of vehicles at an intersection in Seoul, Korea, which showed a dynamic behavior of lane changes near the intersection with different congestion levels.

3.5.4 Effect of traffic heterogeneity

All models described so far are based on the assumption of homogeneous traffic conditions; therefore they are not applicable in conditions of mixed traffic, which is however more likely to occur in real life.

Heavy or high occupancy vehicles operational characteristics affect capacity and Level of Service (LOS) of roadways, particularly at intersections. Lorries, busses and trams need more time to pass than ordinary passenger cars. The headway is larger and thus the saturation flow is lower. In order to simplify calculations, the difference between these vehicles and passenger cars is taken into account by giving them a weight. The adverse effects of a heavy vehicle in a traffic stream are commonly taken care of by converting a truck to a Passenger Car Equivalent (PCE) number. The concept of passenger car equivalent (PCE) is used in Highway Capacity Manual to account for the adverse effects of heavy vehicles and buses on traffic operation at signalized intersections. The HCM

2000 defines the term "passenger car equivalent" (PCE) as "the number of passenger cars that would use the same amount of freeway capacity as one truck/bus or RV under prevailing roadway and traffic conditions".

Traffic composition and vehicle type fractions are usually assumed constant in practical studies. In reality these fractions are random variables. This variability may affect the dynamic and stochastic behavior of overflow queues and delays. Little research has been carried on the impact of traffic heterogeneity in the estimation of queue lengths and delays at signalized intersections. Kang (Kang, 2000) used the microscopic software INTEGRATION (Van Aerde, 2001) to analyze queues, number of stops and delays in relation with the frequency of bus stops and the overall traffic density measured in Personal Car Equivalents (PCE). It turned out that all three measures are non-linearly increasing with the decreasing of bus headways and the increasing of the traffic density. As a consequence, models derived under the assumption of homogeneous traffic cannot be applied in such conditions by simply converting the traffic composition into equivalent homogeneous car traffic with constant rates of conversion.

3.5.5 The uncertainty of delays and queues

Delays that individual vehicles may experience at a signalized intersection are usually subject to large variation due to randomness of traffic arrivals and interruption caused by traffic signal control (Fu, 2000). The assessment of dynamic control schemes would benefit from models of queues and delays, which consider also their variability and time-dependency. For example, having knowledge of the variability of delays makes it possible to estimate the confidence limits about the mean delays and thus provide a more informative comparison of alternative signal plans in identifying optimal signal settings. By considering the variability of delay, more reliable signal control strategies may be generated resulting in improved Level of Service (LOS) of signalized intersections.

Very little consideration has been given to the estimation of the queue variability. Haight (Haight, 1959) firstly derived a probability distribution of the overflow queue length assuming Poisson arrivals and constant headway in the service process. This approach was extended by Mung (Mung, 1996) for a general arrival distribution. Both Haight's and Mung's models are characterized by a high complexity. Newell (Newell, 1971) formulates mathematically this problem using renewal theory (see also Appendix A). This approach inspired the work of Olszewski (Olszewski, 1990a) and other works from the same author (e.g. (Olszewski, 1990b), (Olszewski, 1993) (Olszewski, 1994)), who investigated the queue length distribution in time using a Markov Chain process. He simulated various demand conditions and computed the average and the standard deviation of queues and delays showing a high coefficient of variation especially in conditions near capacity, which is the target of most of the optimal traffic control schemes. Heidemann (Heidemann, 1994), inspired by the work of Meissl (Meissl, 1963), developed analytically the probability generating function of the queue length distribution assuming Poisson arrival process and fixed time control, but this method can only be solved numerically. Fu and Hellinga (Fu, 2000) firstly developed an approximate model for the variance of delays in time assuming stationary demand rate:

$$\sigma^2\{W(T)\} = \frac{t_C{}^2 \cdot (1-y)^3 \cdot (1+3y-4y\cdot\min(1,x))}{12\cdot(1-y\cdot\min(1,x))} + \ldots$$
$$+\left\{\frac{I\cdot T\cdot x}{2\cdot t_C} + \frac{T^2\cdot(1-\max(1,x))^2}{12}\right\}\cdot e^{-\left(\frac{x_0}{x}\right)^{\beta}} \quad (3.35)$$

where $y = q/s$ is the flow-to-saturation flow ratio. The model has been compared with Markov Chain results and with simulation demonstrating very good agreement among them.

There is no formulation however for the variability of queues. The knowledge of the overflow queue variability (e.g. by knowing the standard deviation, or the coefficient of variance) with respect to the variability of the demand and the supply systems can also add a physical meaning to the computed queue length. A stochastic modeling approach for example may be appropriate to characterize the queue states with a known statistical distribution and compute the risk of designing an infrastructure according to expected queues and delays, or with a fixed percentile. This information can be valuable if one wants to compute more accurately the expected costs when for example spillback is likely to occur. Chapter 8 discusses this issue in more detail.

3.6 Extensions for application in general networks

3.6.1 Effect of upstream signals

An isolated intersection works in general very differently from an intersection within an arterial street. In fact, the inflow on one intersection somewhere in an arterial road is limited by the capacities and green times of the different links on the upstream intersection. On the other hand, traffic signals placed in an arterial corridor can be coordinated with one another in order to maximize the likelihood that vehicles passing an intersection upstream will arrive at each intersection downstream during the green phase, optimizing the total throughput of the system. Therefore, the effect of an upstream signal on the operation of all subsequent downstream signals can be divided into two phenomena: 1) platooning effect and 2) filtering effect.

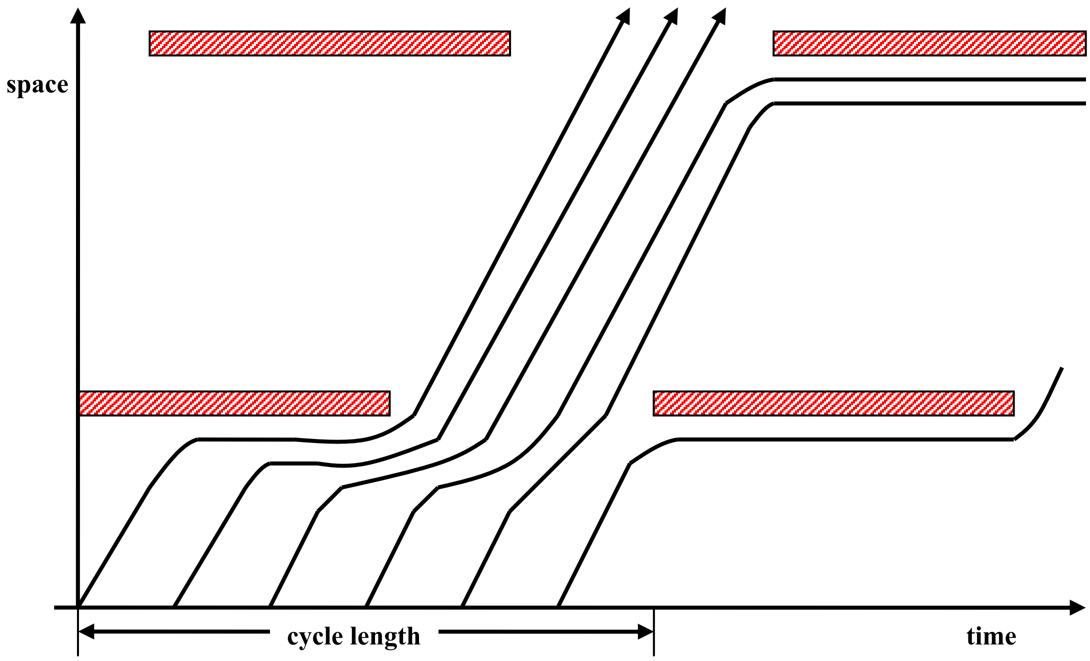

Figure 3.5: Time-space diagram and trajectory of vehicles at paired intersections

Figure 3.5 schematizes the two effects by displaying the trajectory of vehicles in a time-space diagram. Vehicles are assumed in the graph to arrive uniformly at the first signal during a cycle, some vehicles arrive during the red phase and will pass the intersection only after the signal turns green and the preceding vehicles in the queue have been served. Some vehicles arriving during the green phase need to stop or at least to decelerate if at the moment of arrival there are still vehicles in the queue that need to be served, some others may not have time to be served within the green time because of this queue, or because they arrive late at the service point. Vehicles headways are smaller after the first signal, and if the distance among signals is small enough, they will still be bunched when arriving at the second signal (platooning). Moreover the first signal reduces the number of vehicles arriving at the second signal (filtering).

The platooning effect is related to the arrival distribution of vehicles within a cycle, thus it influences primarily the uniform delay component. The second concerns the variability of arrivals within the cycle, and in particular the maximum number of vehicles that can pass an intersection during the green phase. This influences mainly the stochastic and the overflow components.

3.6.1.1 Platooning effect

The signal operation produces the effect of bunching vehicles arriving during one cycle, resulting in non-uniform headways for the departing ones, as seen in Figure 3.5. This effect can be exploited by choosing a proper time-offset among green phases of the intersections constituting the arterial corridor. If then two closely spaced signals are properly coordinated, all vehicles leaving an intersection upstream can hypothetically be served within one green phase and no arrivals will be observed during the red phase. The way signals are coordinated influences the probability that vehicles passing an intersection upstream will arrive at a signal downstream during the green phase. This effect has been modeled by including a multiplicative factor to the uniform delay component (*progression adjustment factor*).

All capacity manuals described in Section 3.3.4 agree in expressing this factor as the following formula:

$$PF = \frac{(1 - P_{coord}) \cdot f_p}{1 - t_{g,u} / t_{C,u}} \tag{3.36}$$

The uniform component is therefore decreasing with the increasing probability P_{coord} of vehicles arriving during the green phase, with the ration between green time and cycle time $t_{g,u} / t_{C,u}$ at the upstream intersection and with an adjustment factor f_p, which depends on green-to-capacity ratio and the shape of the arrival profile. The probability P is estimated in the model using field data or making assumptions on the arrival profile. In reality this value is variable due to the way traffic propagates randomly in between the two intersections. Models of platoon dispersion have been given in the past in order to mathematically estimate the expectation value of the probability P, especially in function of the signal distance and the average speed in the section. In the Highway Capacity Manual, one random and five non-random arrival types are identified. HCM-estimated delays were found to be significantly different from field-measured delays by Benekohal et al. (Benekohal, 1999).

The bunching property of signals diminishes with the distance among signals due to the variability of vehicle behavior. This phenomenon is usually referred to as *platoon diffusion* or *dispersion*. Pacey (Pacey, 1956) proposed a travel time distribution function for the propagation of vehicles along a road with unrestricted overtaking assuming normally distributed speeds. Hillier and Rothery (Hillier, 1967) showed the diffusion phenomenon using field data and the distance-dependency of this phenomenon. They concluded that the signal offset influences the uniform delay component, while it does not influence consistently the overflow delay component.

3.6.1.2 Filtering effect

If the impact of vehicles entering the system from secondary roads or other streams than the arterial roads can be neglected, the maximum number of vehicles approaching the downstream intersection will be somewhat related to the capacity of the upstream one. Therefore, the variability of arrivals and the maximum number of vehicles observed at downstream intersections is strongly influenced by the metering property of upstream signals.

Newell (Newell, 1990), Olszewski (Olszewski, 1990a) and Van As (Van As, 1991) contributed with the first theoretical studies. Newell studied an idealized arterial network with no turning traffic. The system was also considered in steady state equilibrium. Based on these assumptions, Newell concluded that the overflow delay depends only on the critical intersection. He assumed that in the corridor there is always one intersection among the others that works as bottleneck. If the signals are perfectly coordinated and equally set the first intersection will work as a filter. This means that the overflow queue can be observed only in one intersection while in the others the stochastic component can always be neglected. The author concludes that, even with the best settings possible, the overflow queue will be simply distributed along the signals and not reduced. Olszewski confirmed these conclusions using a Markov Chain approach. Van As observed, using real data collected in South Africa, that all the available models at that time highly overestimated the overflow queue by assuming all intersections as isolated, confirming the conclusions given by Newell and Olszewski. The signal optimization program TRANSYT considers every intersection as isolated for the calculation of the random delay, regardless of filtering by upstream sections.

Using a Markov Chain approach, Van As (Van As, 1991) developed also an approximate expression for the coefficient of variance allowing one to calculate in series the variances of all downstream intersections known the arrival distribution at the first signal. He provided an approximate formula to transform the dispersion index of arrivals I_a, at the upstream intersection into the dispersion index of departures B:

$$B = I_a \cdot \exp(-1.3 \cdot F^{0.627}) \tag{3.37}$$

with

$$F = \frac{Q_O}{\sqrt{I_a \cdot q \cdot c}} \tag{3.38}$$

Tarko et al. (Tarko, 1993) studied a system with two intersections proposing two ways of correcting the delay formulae to account for the filtering effect. The practitioners can both reduce the departures from the upstream intersection or increase the capacity of the downstream one. The authors propose also to extend these modifications to the time-dependent queue formula. Tarko and Rouphail (Tarko, 1995) studied an arterial network including turning and merging streams. The authors highlight three factors that have to be included in the calculation of the queues in an arterial network: merging, splitting and filtering effects. Using statistical analysis they confirmed that from the practical viewpoint Newell's approximation method or the simple formula provided by Van As may be sufficient.

3.6.2 Effect of dynamic controllers

Signal modes of operation can be distinguished into three main categories (USDOT, 1996):

- *Fixed and pre-timed controllers*: where the structure and timing of the traffic control process are determined in advance;
- *Actuated controllers*: where individual vehicles are detected and the information from detectors is used to influence the structure and timing of the control program;
- *Adaptive controllers*: where information about the whole traffic situation is used to take decisions about the progress of the control program.

The previous sections have exclusively dealt with models derived under the assumption of fixed time control.

For pre-timed control the information about the traffic situation is used to develop a structure of the control program and to determine also the length of each phase. That means that a fixed time control program is only suited for the traffic situation for which it has been designed. Moreover, even the volumes for which the fixed time program has been designed are average volumes. In each cycle the arrivals will have random variations around the average pattern. Adapting traffic control to the actual traffic situation with traffic responsive controls like actuated and adaptive systems can reduce these delays due to the demand fluctuations.

An actuated controller operates signals according to actual arrival of vehicles at the intersection. Green times and cycle times are determined by the number of vehicles arriving at the intersection during the red phase and their headway distribution during the green phase. Main difference with the adaptive controllers is that actuated controllers do not attempt to optimize traffic by means of e.g. total delay, number of stops, queue length etc. Adaptive signal systems monitor the traffic situation in real-time and adapt the signal settings in order to optimize the traffic operations.

The different logic framing the two systems reflects in a different delay and queue length estimation problem. The various models in literature have been described in the remaining of the section.

3.6.2.1 Traffic actuated signals

The large number of vehicle actuated control types designed in the past and the complex architecture of such systems justifies the lack of a general theory of expectation values of queues and delays with traffic actuated controls (Rouphail, 2000). The first model of delay with a simple traffic actuated signal was proposed by Morris and Pak-Poy (Morris, 1967) with an application on a signal coordinating two one-way streets. For each traffic condition they computed the optimal vehicle interval to minimize the total delay. Newell (Newell, 1971) studied the same problem under the assumption of stationary arrival process in undersaturated conditions but near capacity. He developed approximate formulas of delays and expectation values of green and red extensions in function of the average arrivals at fully-actuated signals. Dunne (Dunne, 1967) proposed a delay model derived by assuming a Binomial arrival process and green times determined by the queue length detected. Cowan (Cowan, 1978) used bunched exponential distribution for the arrivals to compute expected green, red times and delay. Webster's (Webster, 1958) delay Formula (3.8) was adapted by Courage and Papapanou (Courage, 1977) to compute the average cycle length with actuated signals. Optimal cycle lengths were used to compute pre-timed controls using the formula:

$$t_C^* = \frac{1.5 \cdot L + 5}{1 - \sum_i y_i} \tag{3.39}$$

where L is the total lost time in the cycle and y_i is the volume to saturation flow ratio. Fully-actuated signals are instead computed considering average cycle length. The Highway Capacity Manual (TRB, 2000) considers a discount factor of 0.85 to multiply the uniform delay component. The manual also gives an approximate expression of the average signal cycle:

$$t_C{}^* = \frac{x_c \cdot L}{x_c - \sum_i y_i} \tag{3.40}$$

where x_c is the critical volume to capacity ratio. The effective green is given by:

$$t_{g,i}^* = \frac{y_i}{x_i} \cdot t_C{}^* \tag{3.41}$$

where x_i and y_i are respectively the volume-to-capacity and the volume-to-saturation flow ratios for the approach *i*.

Lin and Mazdeysa (Lin, 1983) proposed an extension of the Webster's delay formula, using the above expressions for the green and cycle times, by including two extra parameters, K_1 and K_2, in the form:

$$W = 0.9\left\{\frac{t_C^*(1-K_1\cdot t_g^*/t_C^*)^2}{2\cdot[1-K_1\cdot(t_g^*/t_C^*)\cdot K_2\cdot x]}+\frac{3600\cdot K_2\cdot x^2}{2\cdot q\cdot(1-K_2\cdot x)}\right\} \qquad (3.42)$$

Li et al. (Li, 1994) proposed to include a parameter k in the Australian Capacity Manual time-dependent overflow delay model to account for fully actuated signals:

$$W(T) = 900\cdot T\cdot\left(x-1+\sqrt{(x-1)^2+\frac{8\cdot k\cdot x}{cT}}\right) \qquad (3.60)$$

Using this formulation, Akcelik et al. (Akcelik, 1997) developed simple heuristic formulas to estimate average cycle times and green times for various types of actuated signal control. This method was successfully tested in US and adopted in the 1997 version of the Highway Capacity Manual (TRB, 2000). Analytical models for delay, queue length, clearance time etc. were also developed and validated by comparison with microsimulation.

Although several delay formulas have been proposed to account for traffic actuated control the complex behavior of such system needs still to be fully explored in a dynamic environment. This will be the topic of Chapter 8 of this thesis.

3.6.2.2 Adaptive signals

The development of adaptive control systems and algorithms is increasing with the advances in communication networks, computer processing speed and sensor technologies. The principle of adaptive control was first developed by Miller (Miller, 1963), who proposed an optimization algorithm based on an online traffic model. The first implementation in practice started few years later with the model PLIDENT (Holroyd, 1971), but it did not succeed according to the expectations. Hunt et al. (Hunt, 1982) developed the model SCOOT, which minimizes the total delay by the smooth adaptation of splits, cycle times and offsets. Another network control system was developed and implemented in Australia under the name SCATS (Lowrie, 1982). The two models had several successful implementations reporting an average 10-12% reduction of the total delay with respect to optimal pre-timed controls (Boillot, 1992). SCOOT and SCATS are still the most adopted control methods with 170 implementations all over the world (Friedrich, 2002). Successful implementation in practice has been reported and confirmed by several applied studies after SCOOT, e.g.

MOVA in U.K. (Vincent, 1988), PRODYN in France (Henry, 1983), aaSIDRA in Australia (Akcelik, 2002) and OPAC in the U.S. (Gartner, 1982).

The large number of implementations justifies the numerous assessment studies based on field data. On the other hand the traffic models employed are, due to online requirements, rather simple. Although the development of adaptive control algorithms are increasing and are consistently preferred to the traffic actuated ones, the research on queue and delay estimation with such control strategies is very limited. Brookes and Bell (Brookes, 1991) used Markov Chains to compute the expected queues, delays and stops at adaptive signals. Optimal settings were computed using three heuristic approaches, while delays were computed in sequence using a rolling horizon approach.

Despite the large improvement reported by these controls with respect to the pre-phased signal plans, these improvements are still limited by the lack of an accurate prediction of traffic demands over the projected time horizon.

3.7 Modeling queue and delay dynamics using simulation

The vast set of models presented so far indicates that there is still no clear insight into the way delays are experienced by the drivers under the various conditions of traffic at signalized intersections. There is no general formula which could encompass all aspects of traffic heterogeneity and propagation across a signal. Although several modifications of the queue and delay models to account for all these aspects are still a subject of research, they still can represent real life up to a certain degree. This section discusses the alternative approach to the above analytic models represented by microscopic simulation programs. Since this thesis is focused primarily on analytic models, only a limited description of simulation methods is given. Moreover each simulation program is characterized by different modeling features, justifying the development of a large number of simulation software packets. For a more detailed explanation of how these systems work in each program one can refer to their user’s manual.

Microscopic simulation models represent a way to estimate traffic conditions in a realistic way since they focus on the behavior of each single vehicle and the interactions with other vehicles in the system. The characteristics of each element and their interactions are modeled using still mathematical relationships, which represent, as accurately as possible, real world systems.

The complexity and level of detail of such models is continuously increasing according to the increasing power of computers. Microsimulation models reach nowadays levels of detail, which are not expected to be ever reachable by analytic models. In fact, the supply characteristics represented in delay formulas by means of aggregated measures, like the capacity, are derived in these programs by specifying the physical configuration of the network (e.g. length of the sections, width of the lanes etc.). The flow propagation is on the other hand simulated through an assigned demand, which controls the distribution of vehicle headways at the moment each vehicle is loaded in the system. The position of each vehicle is tracked all along the network taking in consideration its characteristics (e.g. dimensions of the vehicle, desired speed, route choice etc.) and its interaction with the supply system and the other vehicles, which influence their free driving behavior (gap acceptance, car-following behavior etc.). This property allows one to use simulation programs successfully in a variety of transportation problems. Here are listed some examples:

- When mathematical models do not catch the complexity of a process or to investigate the validity of such mathematical models;
- To represent dynamic environments and analyze at the vehicle level some phenomena, which are difficult to understand at the aggregate level;
- To evaluate the performance and compare different management strategies;
- To test new infrastructures;
- To operate safety and impact analyses;

On the other hand, these models are characterized by a complex architecture and a relatively long processing time compared to analytic models. This makes some tasks difficult, like solving iterative processes (like to solve traffic assignment processes or optimizations), and it justifies the appreciation still given to analytical approaches.

However, microscopic simulation programs represent undoubtedly valuable tools to represent the traffic flow dynamics in a very realistic way. The complex behavior of traffic flows at signalized intersections has justified the large use of microscopic programs for analyzing the performance under different scenarios. The opportunity to control the input variables (e.g. by loading stationary or non-stationary demand patterns) allows easy comparison with other models that are characterized by some simplifying assumptions. This property is confirmed by their frequent use in calibration and validation of macroscopic and mesoscopic models.

For example Dion et al. (Dion, 2005) used the microsimulation program INTEGRATION (Van Aerde, 2001) to validate and compare different analytic approaches (vertical and horizontal queuing theory, and capacity guides) and they investigated the behavior of the standard deviation and the coefficient of variation of delays in a one lane road with fixed signal, showing the variability of results to be the highest when volumes to capacity ratios are near the unity. This approach is also adopted in chapter 7 of this thesis.

It is certain a limitation of a model to be validated and calibrated using another model, although more refined. Validation using field data represent an indispensable and irreplaceable step. However, complex dynamic and stochastic systems like the traffic control process are sensitive to many state and control variables (Chapter 2 has shown the various sources of variability, which determine this process). It is difficult, if not very unlikely, that real traffic can be observed with e.g. long periods of stationary conditions. Alternative could be for example simulating this experiment with real cars, but this experiment is definitely too expensive and time consuming. The advancement of new data collection methods and monitoring tools may in the future overcome this limitation.

3.8 Summary

Traffic control is designed to guarantee safety conditions for the traffic operations at intersections while keeping reasonable waiting times for the travelers. Nevertheless, the delay at signalized intersections represents the main component of the total delay experienced by the urban road users. For this reason great importance has been given in the last 50 years to the development of delay models. Although the research has spent large efforts in deriving models, which could explain the complexity of the traffic operations at signal controls, it seems that little progress has been done so far in the understanding of the dynamic and the stochastic behavior of the traffic flow at these sections.

This chapter gives three main contributions:

1. It describes thoroughly the state-of-art and practice of analytic queue length and delay estimation models at signalized intersections. Theoretical and approximate approaches have been described for both steady state and dynamic conditions, as well as for isolated intersections and arterial corridors, and for fixed-timed and time-dependent control schemes.

2. It underlines the drawbacks of using the available analytic formulas, especially in the modeling of the variability of traffic. The complex behavior of queues at signals and the large variety of cases one can observe in real life limits the validity and the applicability of such models.

3. It discusses the delay modeling issue by stressing the various assumptions underlining the models presented in this state-of-the-art chapter, e.g. signals within a network, multiple lanes, dynamic controls etc.

It seems that no model is able to describe the queuing process in a continuous way. There is still discrepancy between static models, which include a stochastic component due to the variability of the arrival process, and the dynamic models. A smooth transition of the stochastic term is not yet caught in one exact formulation. Moreover, the available models are not able to describe both the increasing and the decreasing phases of a queuing process. This limits their application in e.g. peak-hour analysis, or dynamic network loading processes.

An alternative to analytic approaches is using simulation, which allows the estimation of delays at the vehicle level and represent the network in a very realistic way. On the other hand, some transportation problems involving optimization algorithms or any iterative process requires fast travel time models, and microscopic simulation is still not a competitive solution. Solution can lie in between; mesoscopic models can represent in some cases valid alternatives to macroscopic and microscopic models, and they can allow the analyst to evaluate travel times in a stochastic fashion.

Chapter 4 will describe a mesoscopic approach to the queue length and delay estimation problem in a single lane isolated intersection, and it will show the opportunity to use this approach to travel time estimation problems in general traffic systems. Chapter 7 shows some extensions to the simple problem presented in Chapter 4 in cases of arterial corridors, multilane sections and time-dependent control.

4

Probabilistic formulation of queues and delays at signals

4.1 Introduction

Chapter 3 has described the state-of-the-art of modeling control delays at signalized intersections. The calculation of delays at an intersection is an old problem that has been studied and solved by many researchers in the last 50 years since the work of Beckmann (Beckmann 1956). Exact formulations for steady-state conditions as much as exact formulations for the delay in oversaturated conditions have been developed in the past, but no smooth transition is modeled in between these two states using a theoretical approach. In fact only heuristics have been used to solve this gap. It has been the merit of researchers like Kimber and Hollis (Kimber 1979) and Akcelik (Akcelik 1980) to solve this issue by introducing a kind of transformation that gives a smooth transition for the delay expressions for undersaturated and oversaturated conditions. The approach that is followed is a heuristic one, i.e. the formula that is derived fits with the asymptotic situations for very low and very high saturation of the intersection and in the regime of nearly saturated or slightly oversaturated intersections the length of the overflow queue is obtained by some interpolation approach.

The main source of error lies in the way past models deal with the dynamic and the stochastic character of overflow queues, which are assumed in steady-state models to be in equilibrium during the whole evaluation period, while it has a time-dependent expression only during oversaturated conditions. In reality a time-dependency is observable also in conditions of average arrivals below the capacity of the signal, as Chapter 4 shows by computing the probability distribution of the overflow queue length as a Markov Chain process.

As already shown in past studies, the cycle-by-cycle Markov Chain model enables one to consider the effects of the stochastic nature of the arrival and the service processes in the dynamics of queues and delays regardless of what happens within a cycle. This methodology is however very useful to evaluate the time-dependent profile of these measures and their uncertainty but yet no information is given on what happens within the cycle, i.e. how the variable arrivals and departures affect the way queues form and dissolve within the cycle.

This chapter reconsiders the problem by using a Markov chain model for the probability distribution of queue length at any point in time. The discrete-time formulation proposed by previous studies (as introduced in Section 3.5.1) and presented in the following of this chapter is extended to a continuous-time formulation. From the dynamics of the expectation value of the queue length, a formula for the delay in fixed time traffic control is derived. The effect of the overflow queue is therefore made explicit also within the cycle. The result is a new formula and, even more important, a clearer understanding of the role of the overflow queue in the control delay at a signal.

This chapter is structured as follows. Section 4.2 presents the probabilistic model formulation for the queuing process at fixed time signals in discrete time. Section 4.3 applies this methodology to the computation of the control delay in a cycle. Section 4.4 compares the results of the newly developed model with well-known delay formulas presented in Chapter 3. Finally Section 4.5 summarizes the contributions of this chapter.

4.2 The Markov chain process

The process of traffic arriving and leaving at signalized intersections has a stochastic character; in particular, the queue length dynamics is the result of a stochastic process. This process is defined in its cycle-to-cycle case to be a temporal sequence of stochastic variables $\{Q_t\} = \{Q_0, Q_1, Q_2, ...\}$. The stochastic nature of this system derives from two sources: the input demand, mainly its composition and quantity, and the service mechanism or supply, namely the number of vehicles served within a cycle. An analytical representation of the signal system is formally equal to Catling's deterministic expression (Catling 1977):

$$Q_{t+1} = \max\{Q_t + q - s, 0\} \tag{4.1}$$

where q is the number of arrivals and s is the number of departures during $[t, t+1]$. What differs from the deterministic case is that all components of this expression are stochastic variables, thus represented by probability distributions. Consequently, also the queue length at a certain time is a realization from a probability distribution.

The signalized intersection system is governed by a cyclic mechanism. This allows one to use discrete time steps, equal to the cycle length, instead of the continuous approach, if one is interested only in computing the overflow queue length distribution at the end of the green phases. If cycle lengths are invariant with time, for example in the case of fixed time controls, the time steps are usually equal in length, while in cases of time-dependent control schemes these time steps are variable. Thus, accounting again for its cyclic property and for the independence of the input variables demand and supply, a state Q_{t+1} is described only by the previous state Q_t and the number of arrivals and departures during the interval $[t, t+1]$, according to Formula (4.1). These assumptions allow the application of a standard method of renewal theory, the Markov Chains (Markov 1971), to solve analytically this problem and to compute a complete probability distribution of queues in time.

Markov Chain models have already been applied to describe time-dependent processes in transportation, both in the signalized intersection context and in other contexts. Cronje (Cronje 1983) analyzed existing formulas, namely, Webster's (Webster 1958) and Miller's (Miller 1968) equations for average delay, overflow, and average number of stops for under-saturated conditions using a Markov Chains using a geometric probability distribution for the arrivals. The properties of the geometric probability distribution were applied to the equation to obtain a simple equation, thus reducing computing time. Olszewski ((Olszewski 1990), (Olszewski 1990), (Olszewski 1994)) used this technique to analyze the queue and delay probability distributions in time especially to consider the effect of non-zero initial queue and non-uniform arrivals.

The convenience of applying this technique to simulate the possible traffic states in a stochastic environment is therefore not new in traffic flow theory. However, few studies have uncovered the opportunity to use this method to derive analytical expressions suited for planning and design purposes (e.g. (Brilon 1990), (Fu 2000)). The representation of a dynamic process using a Markov Chain approach can be a valid alternative to simulation programs and it can provide a sufficiently large dataset to obtain smooth representations of the expectation value of the queue and its variability and to derive analytical expression for both characteristics in time. The next section formulates the Markov Chain process applied to isolated signalized intersections, which will provide the dataset for the derivation of analytical formulas for the expected value and the standard deviation of the queue length in time developed in Chapter 6.

4.2.1 The isolated signalized intersection

Given the analogous properties of Markov chain and queuing system processes, the dynamics of the queue length and its distribution can be computed with discrete time steps by simple matrix multiplications. Van Zuylen (Van Zuylen 1985) firstly described a Markov model for queues at isolated intersections assuming Poisson arrivals and normally distributed saturation flows. Olszewski (Olszewski 1990) independently developed the idea of applying the Markov chain technique to signal control problems. He showed the different behavior expected values of queues have if a different initial value is assigned together with a constant demand during the whole evaluation period. Variable demand conditions were later analyzed by assuming stepwise constant demand (Olszewski 1990). Following the approach adopted by van Zuylen and Olszewski, a similar Markov model is developed in this chapter to compute the distribution and its evolution in time of the queue length at the end of the green phase. The model computes the queue length distribution under the whole range of demand conditions, providing the dynamic behavior of queues both in undersaturated and oversaturated cases and with time-varying average arrivals. Expected value and standard deviation are derived from the computed probability distribution at each time step, capturing the transient behavior towards the equilibrium value in undersaturated cases, and the linear, deterministic behavior in highly oversaturated cases.

Next section points out the assumptions of the model and the hypotheses gradually introduced throughout the chapter. Firstly the model is shown in its simplest case, namely isolated, single lane intersection with homogeneous traffic composition and fixed time traffic control. Analysis of average conditions and variability is presented under broad conditions of traffic.

4.2.2 Model assumptions

The Markov process requires the specification of the input demand and service rates within a cycle in terms of probability distribution. Beckmann (Beckmann 1956) derived a first expression of expected delay at fixed time signals assuming binomial arrival process, while Dunne (Dunne 1967) used the same assumption to derive an expression of delays with traffic actuated control. Apart from these exceptions, it is a generally accepted hypothesis to consider the arrivals at an isolated intersection within a cycle following a Poisson process (Kang 2000). In practice, the demand is subdivided into periods of stationary conditions, in which the average arrivals do not change significantly from each other. According to the definition of Poisson distribution, this average value represents the only parameter, which defines its shape.

The parameter of the Poisson distribution is assumed constant or stepwise constant within the period of analysis. This assumption allows one to analyze the dynamics in the simplest scenario and to specify the direct relationship between demand and queue evolution. Behavior towards and equilibrium value in undersaturated cases or behavior in oversaturated cases can be studied under this assumption. The hypothesis of Poisson distribution for the arrivals, on the other hand, determines the statistical behavior of the overflow queues computed in this chapter; it is expected that a different distribution will lead to different results. The Markov model can be numerically evaluated using any probability distribution. A distribution function can be characterized for example by collecting a sufficient amount of traffic counts under the same prevailing conditions and finding the most appropriate function to fit these real observations.

It is assumed throughout this thesis that the flow can exceed the effective capacity at the stop line but it is still far below the saturation flow. In theory, a Poisson distribution does not consider an upper limit to the number of vehicles approaching the intersection from one of its arms, which is an unrealistic assumption, even if the very little chance given to these outcomes produces negligible effects on the resulting queue length distribution. In practice, the computation algorithm requires a finite input demand. The upper limit can be then fixed to the maximum number of arrivals determined by the capacity of the upstream link. The probability given to higher arrival rates is then assigned to the maximum value. This assumption can represent cases where the link is full, i.e. queues can build up somewhere upstream the intersection independently of the signal process.

The intersection approach is assumed to serve a maximum possible number of vehicles within a cycle. This value can be also derived from the assumed saturation flow. Effective capacity per cycle, s, is either assumed to be constant or to have a Binomial probability distribution. In this chapter this random variable is considered independent of both the queue length observed at the starting of the cycle and the input flow q. This assumption may not be completely correct; different distributions may be observed with different congestion levels, as a result of a different car-following behavior (e.g. the stress for waiting at long queues may reduce the reaction time of drivers or make them accept shorter headways among one another).

Apart from demand and capacity for each cycle, the computation of the queue distribution requires the specification of an initial queue state Q_0. This value can be represented by a deterministic value or also by a probability distribution. Both assumptions might be used in practice. For example a deterministic initial value can be assumed if the queue estimation follows the observation of a queue at a certain point in time and space and a prediction of the evolution from this state is forecasted. If the analyst wants on the other hand to estimate the dynamics and average time to clear the queue during a peak hour, every period of stationary conditions starts with a certain initial state, which is determined by the distribution computed at the previous period.

The opportunity to compute the number of queued vehicles and their variability allows the analyst to compute also the probability that this number exceeds the available road space designed to contain the waiting users. This feature allows the computation of e.g. the chance of observing spillbacks and gridlocks and to include these costs in the computation of the total delay of a road network or to evaluate the road infrastructure design (i.e. exclusive turning lanes). Nevertheless, the assumption of homogeneous car traffic may limit this computation.

The road geometry is considered in this chapter simply consisting of one lane per direction for the sake of simplicity. Moreover, the signal is assumed in this chapter as isolated, i.e. the arrival distribution is invariant with time. The first assumption guarantees that the mass-balance equation is applicable within the lane, and FIFO condition holds. The second assumption implies that no effect of upstream signals is considered, i.e. filtering or platooning effects. Later in this thesis the effect of upstream signals and the multilane cases are discussed, and a Markov model for multiple lanes, which explicitly considers the lane changing behavior of travelers is later developed (Chapter 8).

4.2.3 Description of the system

This section describes the mechanism driving the evolution of queues in a single lane, isolated intersection. The system is defined in queuing theory as M/D/1 in the case of deterministic service and M/M/1 for the stochastic service case. As described in Chapter 3, most analytic time-dependent queue models have been developed under the assumptions of isolated intersection and single service. To compare these models with the Markov model the same assumptions have been used to compute the queue length distribution in time. Figure 4.1 shows an example of this system. Let t_g, t_r and t_C represent respectively the effective green, red and cycle times of the fixed control. Let L be the total length of the road section upstream the signal. If the intersection has large capacity with respect to the demand within a cycle it is very likely that the whole demand is served before the end of the green phase. If the intersection is oversaturated the length of the green phase is insufficient to completely clear the intersection and residual queues will be most likely observed at the end of the green phase (*overflow queues).*

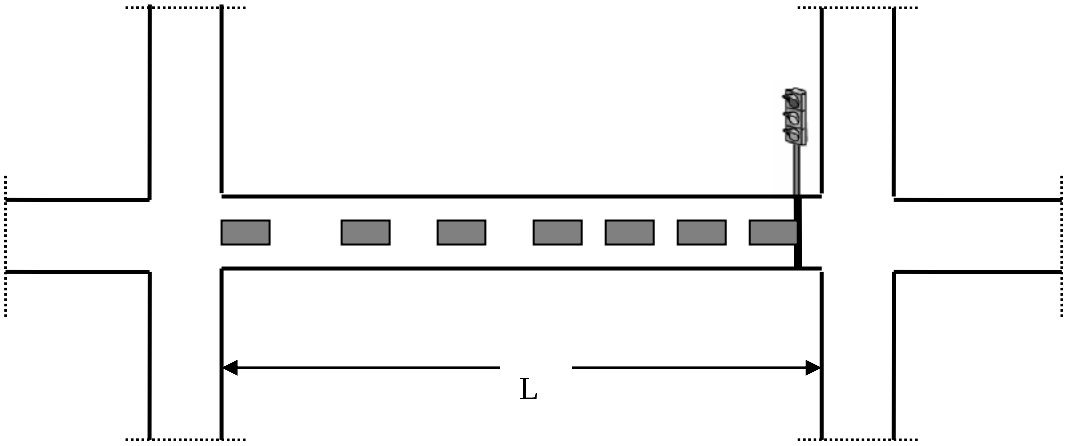

Figure 4.1– Scheme of a single lane intersection

Deterministic models exclude the existence of residual queues if the average volume-to-capacity ratio is less than 1. In reality although for a certain period of time the flow is, in its average, smaller than the capacity, its variability produces a non-zero chance that for some cycles the number of arrivals will be larger than the possible departures, and some vehicles will stop twice or more times before the signal in order to be served. The closer the volume of traffic to the capacity of the signal, the larger the chance to observe these residual queues will be. On the other hand, if the volume-to-capacity ratio is only slightly larger than the unity, there will be still at least for the first cycles a non-zero chance that the arrivals will be less than the departures.

4.3 Overflow queue model formulation

Let $Q_{\max}$ be the maximum value of the queue length, which can be stored in the considered road section, $q_{\max}$ and $d_{\max}$ respectively the maximum number of arrivals and departures possible within a cycle, Formula (4.1) can be computed in a stochastic fashion by first computing the transition matrix Q_{ij}, which represents the probability that the queue length moves from a state j at time $t-1$ to state i at time t. If $j \neq 0$ this probability is expressed by:

$$Q_{ij}(t) = \begin{cases} \Pr(i = j + q_t - d_t) & \forall \quad j \geq i - d_t, \quad q_t \in [0, q_{\max}], \quad d_t \in [0, d_{\max}] \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

Since queues are constrained to be non-negative, when the departures are larger than the sum of the arrivals and the queue at the starting of the cycle, the queue at the end of the green phase will be zero. Obviously, part of this green phase will not be used by any vehicle. According to this consideration the chance of a queue i to become zero is computed with the following condition:

$$Q_{i0}(t) = \begin{cases} \sum_{k=0}^{d_t - i} \Pr(k - q_t = 0) & \forall \quad i \le d_t, \quad q_t \in [0, q_{\max}] \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

If the departures are deterministic, Formula (4.3) computes the probability for a specific queue length j in the transition matrix from each couple (i, a_t). If departures d_t are stochastic, given the range of possible departures $[0, d_{\max}]$ and the assumption of independence of departure and arrival distributions, the transition probability from a state i to a state j is given by:

$$Q_{ij}(t) = \sum_{d_t=0}^{d_{\max}} Q_{ij}(t, d_t) \Pr(d_t) \tag{4.4}$$

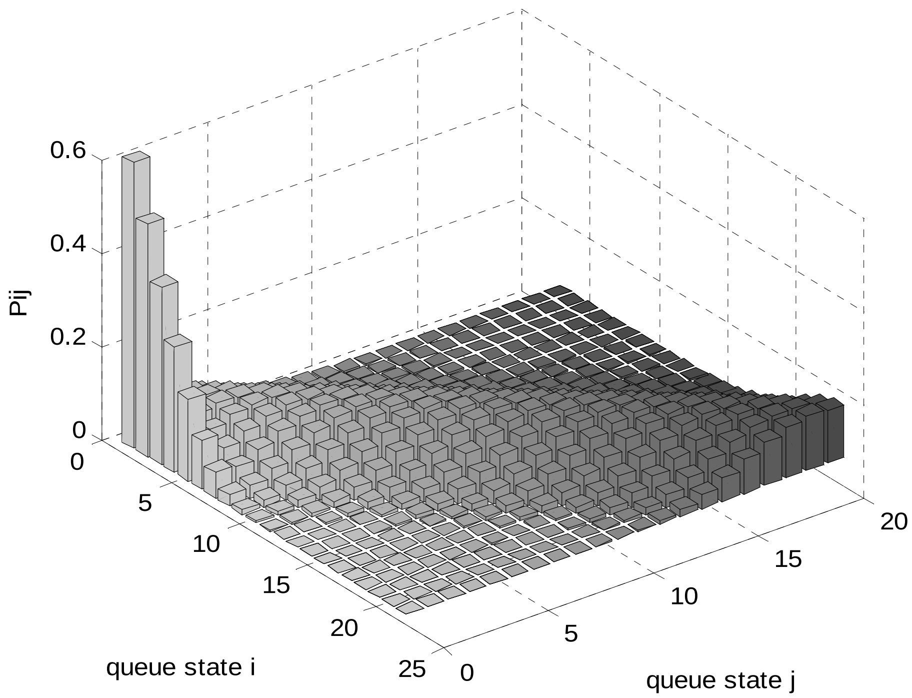

Figure 4.2 – Transition matrix for x=0.975

Figure 4.2 shows an example of queue transition probability for an undersaturated case ($x = 0.95$). If the queue has a non-zero initial state i there is a distribution of probabilities that the queue stays among states adjacent to i. If the queue is zero, there is over 50% chance that it will remain zero also at the following cycle.

Every time step t is uniquely determined once an initial condition Q_0 is assumed. This value, as said, can be a specific value or a stochastic variable. In both conditions the initial condition can be expressed by a vector of initial queue probabilities $\Pr_{Q_0}(0) = \{\Pr_0(0), \Pr_1(0), \Pr_2(0), ..., \Pr_{Q_{\max}}(0)\}$ where the deterministic case can be seen as a special case of this vector where probability is 1 for the deterministic value and zero for the others. Since the queue probability distribution at every time $t-1$ and the transition matrix q_{ij} are, as defined, independent, the probability of each state j is given by:

$$\Pr(Q_o = j, t) = \sum_{i=0}^{Q_{\max}} \Pr(Q_o = i, t-1) \cdot Q_{ij}(t) \tag{4.5}$$

Figure 4.3 displays an example of queue length distribution for a sequence of 20 cycles (x=0,975 and $Q_0 = 10$). The distribution is very flat, whilst the probability of observing a zero queue increases in time. Although it does not reach 10% of the cases it still represents the most likely state. This demonstrates the high variability of such queues and the higher uncertainty on the prediction of the queue length the longer the chosen prediction horizon.

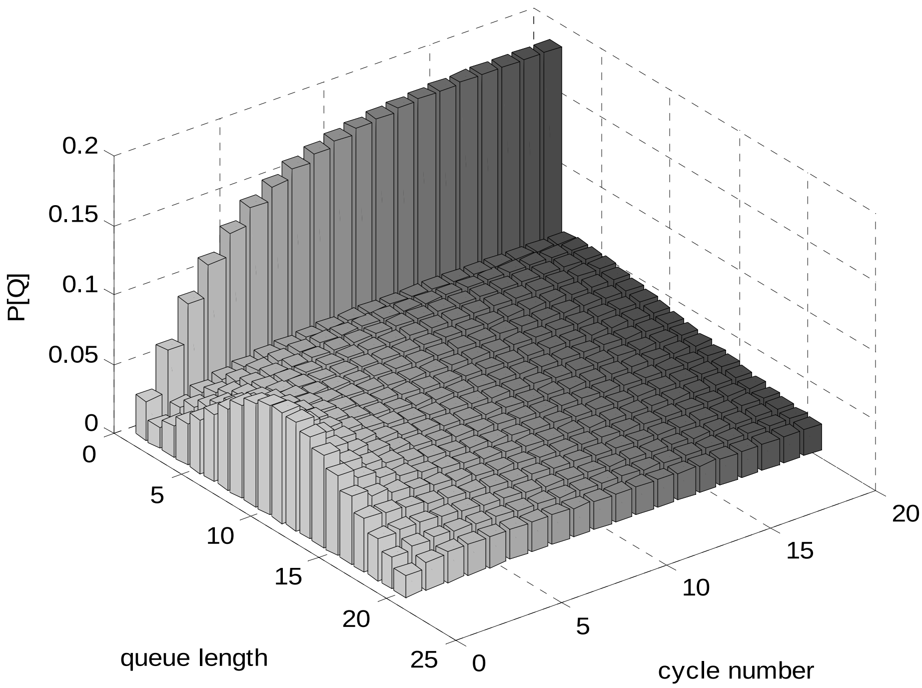

Figure 4.3 - Evolution of queue length probabilities for x=0.95

Expected values and standard deviations are computed with the following equations:

$$E[Q_o(t)] = \sum_{j=0}^{Q_{\max}} j \cdot \Pr(Q_o = j, t) \tag{4.6}$$

$$\sigma\{Q(t)\} = \sqrt{\sum_{j=0}^{Q_{\max}} (j - Q_t)^2 \cdot \Pr(j,t)} \tag{4.7}$$

The next chapter deals with the sensitivity of the above expectation value and the standard deviation of the overflow queue to the assumed stochastic variables (arrivals, departures and initial queue length). The next section describes the relationship between the dynamics of the queue length distribution and the dynamics of the vehicle delay. According to the way queue length has been formulated, delays are also formulated as Markov chain processes.

4.4 Probabilistic formulation of the control delay

Delays are characterized by probability distributions as much as queues, since direct relationship links these two characteristics. This relationship between queue length at each cycle and delay and between their variability was similarly developed by Olszewski (Olszewski 1994). Conventional analytical delay models provide only point estimates of delay, averaged over the period of analysis (typically 15, 30 minutes or 1 hour). Since delays have stochastic and dynamic behavior as much as queues, vehicle delay at each cycle may be consistently different from the computed average. Furthermore, the knowledge of probability distribution of delays in time would allow one to obtain standard deviation and confidence intervals in time, which are valuable input for reliability studies, or for the evaluation of the variability of the expected level-of-service computed when assessing the quality of an intersection or the impact of different control strategies. Travelers might also find this variability useful information to their daily travel decisions.

As reported in Chapter 3, the total vehicle delay can be subdivided into three components (uniform, random and initial queue delay), according to the definition given by the latest version of the Highway Capacity Manual (TRB 2000). The uniform component W_1 reflects the average waiting time of a vehicle approaching the intersection if no queue is present at the intersection. This component represents the probability of a vehicle to arrive at the intersection when the signal is red. All manuals suggest the following formula:

$$W_1 = \frac{t_r^{\ 2}}{2 \cdot (t_C - t_g \cdot \min(x,1))} \tag{4.8}$$

Particular attention is given to a range of volume-to-capacity ratio floating around the unity. In fact, the randomness of vehicle arrivals results in a delay function that tends to a uniform delay model at low v/c ratios and a deterministic over-saturation delay model at high volume-to capacity (v/c) ratios (in excess of 1.3). At v/c ratios in the range of 0.8 to 1.2, the stochastic nature of traffic arrivals results in significantly higher delays than estimated by standard deterministic queuing models. In this range of v/c ratios, the non-linear relationship between delay and the v/c ratio means that the marginal delay associated with an increase in demand is higher than the one associated with a decrease in demand. This causes the delay associated with random arrivals to be higher than the delay associated with uniform arrivals.

4.4.1 Cycle delay model formulation

To compute the overflow delay each vehicle experiences one needs to compute the propagation in time of this delay from the arrival of the vehicle at the intersection till the moment it leaves the intersection, which can happen several cycles onwards.

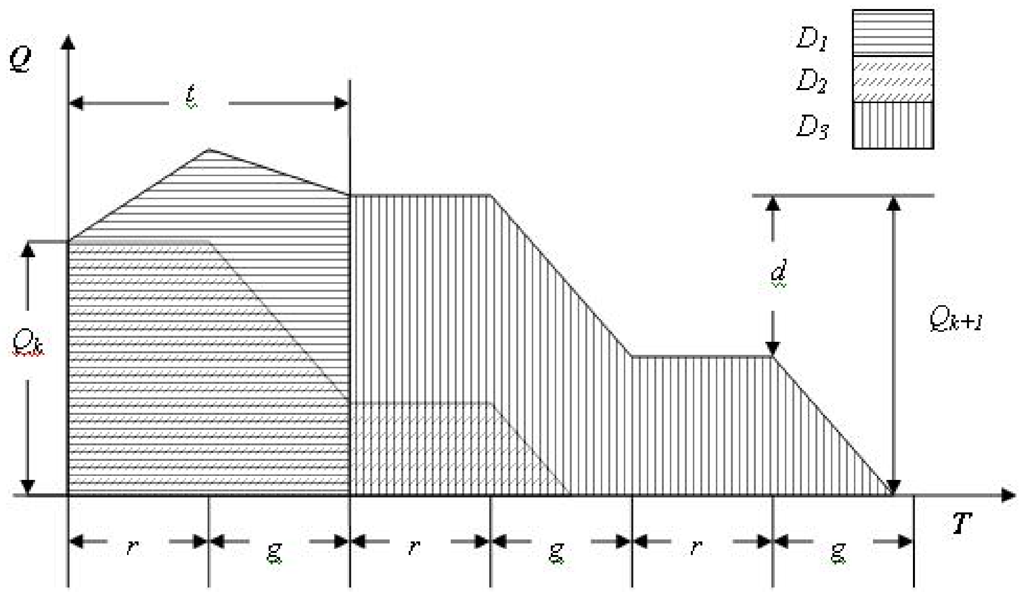

Figure 4.4 – Cycle delay with overflow queue

Figure 4.4 displays the interaction between the three components in a simple schematic example and how a vehicle delay is distributed along adjoining cycles. Vehicles arriving before the vehicle arrives influence its delay only if they are still to be served and thus belong to the queue Q_k . Accordingly, if a vehicle arrives at the approach in later cycles, it will not give any influence as well.

Since a residual queue remains at the start or the end of a cycle, delay at a certain cycle t is caused by vehicles arriving at the intersection during previous periods and during the same cycle t. It may happen that one cycle time is not enough to handle this residual queue and a new residual queue is observed at the end of cycle t. Vehicles arriving at cycle t will then experience a delay that is represented by the following areas showed in the figure:

$$D_t = D_1 - D_2 + D_3 \tag{4.9}$$

The first component D_1 represents the total delay accumulated within cycle k by all vehicles arriving during that cycle and the ones already waiting at the signal. The second term D_2 is the delay experienced by only the vehicles waiting at the signal at the starting of cycle k and finally D_3 is the delay experienced by the vehicles arriving during k and caused by the presence of the residual queue Q_k . If the queue is not completely cleared at the end of a simulation period, one should then take into account that vehicles entering a road section in later periods have an extra delay caused by an overflow queue and they may still accumulate part of their delays after the end of the simulation.

The following method, inspired by the work of Olszewski (Olszewski 1994), provides a systematic computation of vehicle delay for one cycle based on the geometrical relationship drawn in Figure 4.4. The number of departures d is assumed constant throughout the evaluation period. Area D_1 is thus expressed by the following conditions:

$$D_1 = \begin{cases} \dfrac{(2 \cdot Q_t + q_t) \cdot t_C - d \cdot t_g}{2} & \text{if } Q_t + q_t \geq d \\ \dfrac{(2 \cdot Q_t + q_t) \cdot t_C - d \cdot t_g}{2} + \dfrac{(d - Q_t - q_t)^2}{2(d / t_g - q_t / t_C)} & \text{otherwise} \end{cases} \tag{4.10}$$

If t cycles are needed to serve Q vehicles, this last value will decrease cycle by cycle following the sequence $Q, Q-d, Q-2 \cdot d, ..., Q-t \cdot d$ (where t is computed as the smallest integer value such that $Q_t - t \cdot d \leq 0$). Based on this observation the total delay caused by Q vehicles waiting at the start of a cycle t will be computed with the following function:

$$\Psi(Q) = \frac{Q^2 \cdot t_g}{2 \cdot d} + (t+1)\left(Q - \frac{t \cdot d}{2}\right) \cdot t_r \tag{4.11}$$

The first term of (4.11) represents the delay incurred during the green periods while the second is the one incurred during the red periods. Areas D_2 and D_3 are then simply computed by substituting Q respectively with Q_t and Q_{t+1}. The average delay incurred by a vehicle approaching the intersection during cycle t is then computed with the following formula derived from Formulas (4.10) and (4.11):

$$D_t = \frac{D_1 - \Psi(Q_t) + \Psi(Q_{t+1})}{q_t} \tag{4.12}$$

At each cycle, the queue length distribution is computed with the method as described in Section 4.3. The distribution of delays in time will then only depend on the distribution of queues at the start and the end of one cycle and the distribution of the arrivals at that cycle. Since these two characteristics have been assumed independent each other, the probability of each delay is computed by simple multiplications of these two probabilities.

4.5 The within-cycle queuing process

The overflow queue model has shown that the length of the overflow queue can be derived in a fundamental way by describing the probability distribution of the queue length as a Markov chain process in discrete time steps: the probability distribution of the queue length at the end of the green phase depends on the probability distribution of the queue length at the end of the previous green phase and the probability distribution of the arrivals during the cycle (Figure 4.5).

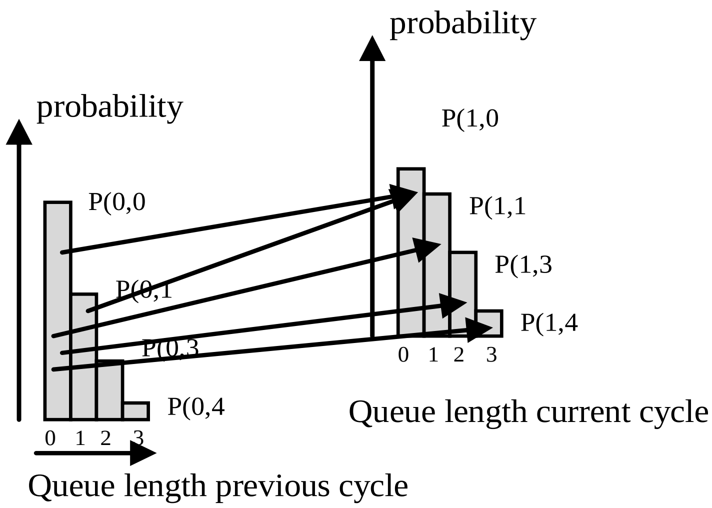

Figure 4. 5 : The transition process of the probability distribution of the queue length at the end of the green phase from cycle to cycle

This relationship has both multiple-to-one and one-to-multiple properties, since each queue state at the previous cycle contributes to the probability of all queue states for the current cycle and, inversely, each queue state at the current cycle is determined by each probability state at the previous cycle. This relationship can be extended to a continuous time step by simply considering green and red phases separately.

If one assumes deterministic departure rate, while no specific distribution is assumed for the arrivals, the transition following the Markov Chain approach in a cycle-to-cycle process can be described with the following general relationship:

$$\Pr\left(Q=j,\tau\right)=\sum_{l=0}^{[j+s\cdot t_g]}\Pr(q_\tau=l,t_C)\cdot\Pr\left(Q=j-l+s\cdot t_g,\tau-1\right) \tag{4.13}$$

Where:

- $\Pr\left(Q_o=j,\tau\right)$ is the probability of a queue length j at the end of the green phase of cycle τ
- $\Pr(q_\tau=l,t_C)$ is the probability of l arrivals in a cycle of length t_C

The square brackets on the summation indicate the integer value of the number. The same approach can be applied for the queue length distribution at any moment during the cycle. Let the signal cycle be divided into two phases, i.e. the effective red and the effective green phase, as described in Chapter 3. During the red phase queues can only grow in time, and their variability is simply determined by the variability of the arriving vehicles. The probability distribution $\Pr(Q=n,t)$ for n vehicles waiting in front of the stop-line at time t and during the red phase can be expressed as:

$$\Pr(Q=n,t)=\sum_{j=0}^{n}\Pr(Q=j,t-\Delta t).\Pr(q_{\Delta t}=n-j,\Delta t) \tag{4.14}$$

Where $\Pr(q_{\Delta t}=n-j,\Delta t)$ is the probability that $n-j$ vehicles arrive in the time between $t-\Delta t$ and t. The expectation value of the queue is then given by:

$$E[Q(t)]=\sum_{n=0}^{\infty}n\cdot\Pr(Q=n,t)=\sum_{n=0}^{\infty}n\sum_{l=0}^{n}\Pr(Q=n-l,t-\Delta t)\cdot\Pr(q_{\Delta t}=l,\Delta t) \tag{4.15}$$

If this expression is rearranged it can be evaluated exactly:

$$E[Q(t)] = \sum_{l=0}^{\infty}\sum_{n=l}^{\infty} n \cdot \Pr(Q = n-l, t-\Delta t) \cdot \Pr(q_{\Delta t} = l, \Delta t) =$$

$$= \sum_{l=0}^{\infty}\sum_{n=l}^{\infty} (n-l) \cdot \Pr(Q = n-l, t-\Delta t) \cdot \Pr(q_{\Delta t} = l, \Delta t) + ... \qquad (4.16)$$

$$+ \sum_{l=0}^{\infty} l \cdot \Pr(q_{\Delta t} = l, \Delta t) \cdot \sum_{n=l}^{\infty} \Pr(Q = n-l, t-\Delta t)$$

Giving the intuitive result:

$$E\left[Q(t)\right] = E\left[Q(t-\Delta t)\right] + E\left[a(\Delta t)\right] = E\left[Q(t-\Delta t)\right] + q \cdot \Delta t \qquad (4.17)$$

Where q is the average arrival rate during Δt. Therefore, the expectation function of the queue length is a simple linear function of the time and grows proportional to the average arrival rate q. This justifies the simplified representation of the queue length as a linear function as is used in many estimations of the delay.

In the green phase in a first approximation, a similar calculation can be made, in which the departures are subtracted from the expectation value of the queue. This means that the expectation value of the queue is a simple linear function as shown in Figure 4.6. Later in this section it is shown that the stochastic behavior of the traffic process causes some deviations of this simple model.

Formula (4.17) expressed the expected queue length as a function of the time during the red phase. Similarly, one can express the queue length probability during the green phase:

$$P(Q = n, t) = \sum_{l=0}^{n+s\cdot\Delta t} \Pr(Q = n - l + s \cdot \Delta t, t - \Delta t).\Pr(q_{\Delta t} = l, \Delta t) \qquad \text{if n>0}$$

$$P(Q = 0, t) = \sum_{l=0}^{[s\Delta t]} \Pr(q_{\Delta t} = l, \Delta t) \sum_{j=0}^{[s\Delta t - l]} P(Q = j, t - \Delta t) \qquad \text{otherwise} \qquad (4.18)$$

Expression (4.18) is just the consequence of the fact that the probability distribution function has only finite values for positive queue lengths. Figure 4.6 shows the evolution process of the distribution function.

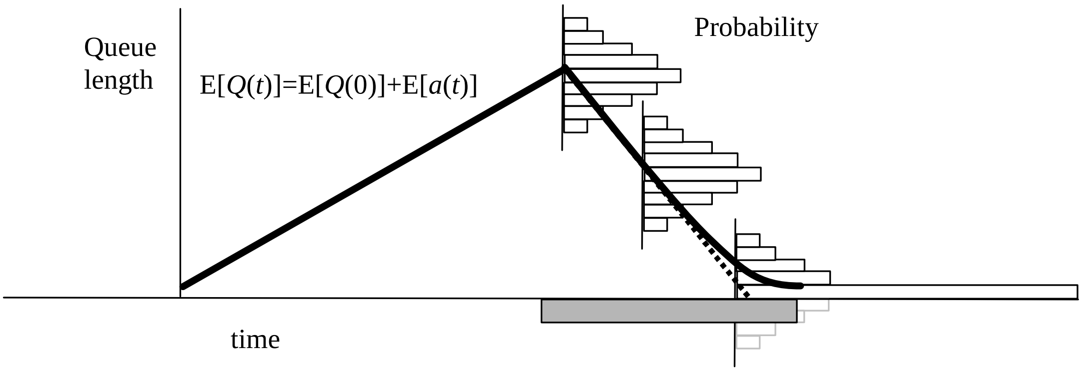

Figure 4.6: The probability distribution of the queue length in the green phase.

As long as the expectation value of the queue length is large, the probability of observing a zero queue during the green phase may be very little. If on the other hand the signal operates near the capacity, this probability can be very high. Comparing this figure with Figure 3.1 the effect of the variability of the arrivals at the end of the green phase is made explicit. The dynamics of the queue in the green phase becomes:

$$E\left[Q(t)\right]=\sum_{n=0}^{\infty}n\cdot\Pr(Q=n,t)=\sum_{n=0}^{\infty}n\sum_{l=0}^{[n+s\cdot\Delta t]}\Pr(Q=n-l+s\cdot\Delta t,t-\Delta t)\cdot\Pr(q_{\Delta t}=l,\Delta t) \quad (4.19)$$

Formula (4.19) shows that the queue initially decreases linearly until the moment that the standard deviation of the queue length distribution becomes of the same order of magnitude as the expectation value of the queue, as shown in figure 4.6.

Equation (4.19) can be rewritten by rearranging the summation:

$$\begin{aligned}
E\left[Q(t)\right]&=\sum_{l=0}^{\infty}\sum_{n=\max(0,[l-s\Delta t])}^{\infty}n\cdot\Pr(Q=n-l+s\cdot\Delta t,t-\Delta t)\cdot\Pr(q_{\Delta t}=l,\Delta t)=\\
&=\sum_{l=0}^{\infty}\Pr(q_{\Delta t}=l,\Delta t)\sum_{n=\max(0,[l-s\Delta t])}^{\infty}(n-l+s\Delta t)\cdot\Pr(Q=n-l+s\Delta t,t-\Delta t)+\\
&+\sum_{l=0}^{\infty}l\cdot\Pr(q_{\Delta t}=l,\Delta t)\sum_{n=\max(0,[l-s\Delta t])}^{\infty}\Pr(Q=n-l+s\Delta t,t-\Delta t)+\\
&-s\cdot\Delta t\sum_{l=0}^{\infty}\Pr(q_{\Delta t}=l,\Delta t)\sum_{n=\max(0,[l-s\Delta t])}^{\infty}\Pr(Q=n-l+s\Delta t,t-\Delta t)
\end{aligned} \quad (4.20)$$

This results in the following formulation:

$$\begin{aligned}
E\left[Q(t)\right]&=E\left[Q(t-\Delta t)\right]+q\cdot\Delta t-s\cdot\Delta t+...\\
&\sum_{l=0}^{[s\cdot\Delta t]}\Pr(q_{\Delta t}=l,\Delta t)\sum_{m=0}^{[s\cdot\Delta t]-l}(s\cdot\Delta t-m-l)\cdot\Pr(Q=m,t-\Delta t)
\end{aligned} \quad (4.21)$$

The first three terms of Equation (4.21) are simply the expression of the linearly decreasing queue, the last term is the effect of the stochastic character of the arrivals. Equation (4.21) can simply be rewritten in terms of the probability distribution function of the queue at $t=0$, the beginning of the cycle, i.e. at the start of the red phase.

$$E\left[Q(t)\right]=E\left[Q(0)\right]+q\cdot t-s\cdot(t-t_r)+\dots \sum_{l=0}^{[s(t-t_r)]}\Pr(q_t=l,t)\sum_{m=0}^{[s(t-t_r)]-l}(s\cdot(t-t_r)-m-l)\cdot P(Q=m,0) \tag{4.22}$$

One may easily verify that the expression approaches zero when $t\to\infty$. The calculation of the third term can be done numerically. If one assumes that the cycle begins with a deterministic state $E\left[Q(0)\right]=Q_0$, the correction term in Equation (4.21) becomes:

$$\sum_{l=0}^{[s\cdot(t-t_r)]}\Pr(q_t=l,t)(s\cdot(t-t_r)-Q_0-l) \tag{4.23}$$

Since the first terms of Equation (4.21), the deterministic queue length, continues also after the moment that the deterministic queue disappears, the correction term has to compensate for that. The deterministic queue is therefore given by the following formulation:

$$Q_{\det}(t)=\begin{cases}Q_0+q\cdot t-s(t-t_r) & \text{if } t<\dfrac{(Q_0+s\cdot t_r)}{s-q}\\ 0 & \text{otherwise}\end{cases} \tag{4.24}$$

Since the first part of Equation (4.21) continues to decrease, the correction term has to compensate the negative part. Figure 4.7 shows the expectation of the queue length computed with Formula (4.21) for different degrees of saturation if a Poisson distribution is assumed for the number of arrivals within the cycle. The results are computed using $t_g=24$, $C=60$, $s=1800$ veh/h and $Q_0=0$. The behavior of the decreasing part of the queue length confirms the illustrative example drawn in Figure 4.6.

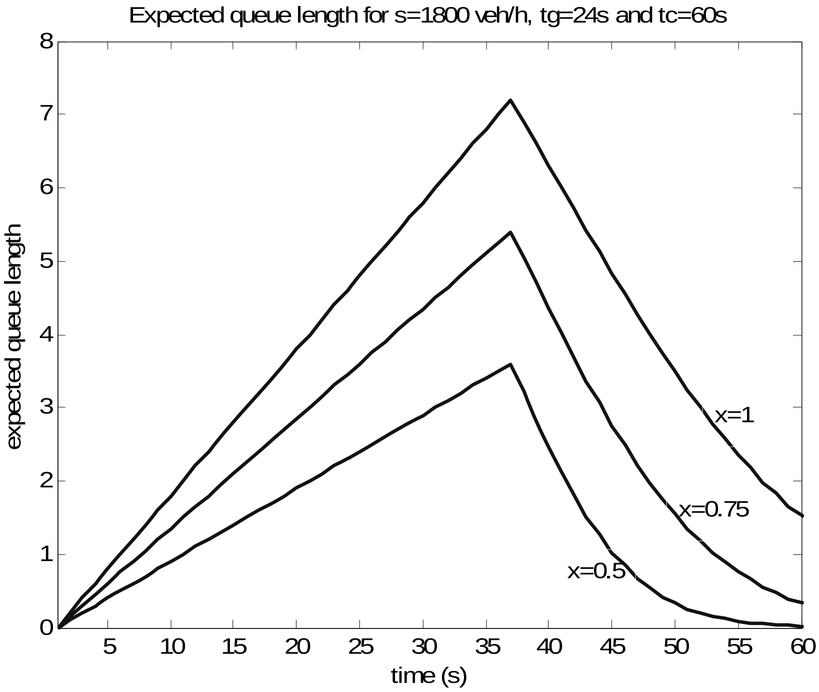

Figure 4.7: Expected queue length within a cycle

The fundamental role of the compensation term is clearly visualized at every demand condition. For very low degrees of saturation (as $x = 0.5$ in the figure) it is small error to consider no expected overflow queue at the end of the cycle. The larger this value, the larger positive expectation value results at the end of the cycle. The compensation term allows one to compute the expectation value also when x=1, cannot be calculated with static models. The new Formula (4.21) enables one to calculate a finite value for the overflow queue at the end of each cycle. This value is nonetheless dependent on the assumed probability distribution. One should expect this value to be different from the one displayed in Figures 4.7 and 4.8 if other distributions than Poisson are assumed.

Figure 4.8 gives also insight into how different cycle lengths affect the expectation value of the queue at the end of the cycle. Although the increase is relatively small in the example, the larger the cycle length, the higher overflow queue is observed at the end of the cycle.

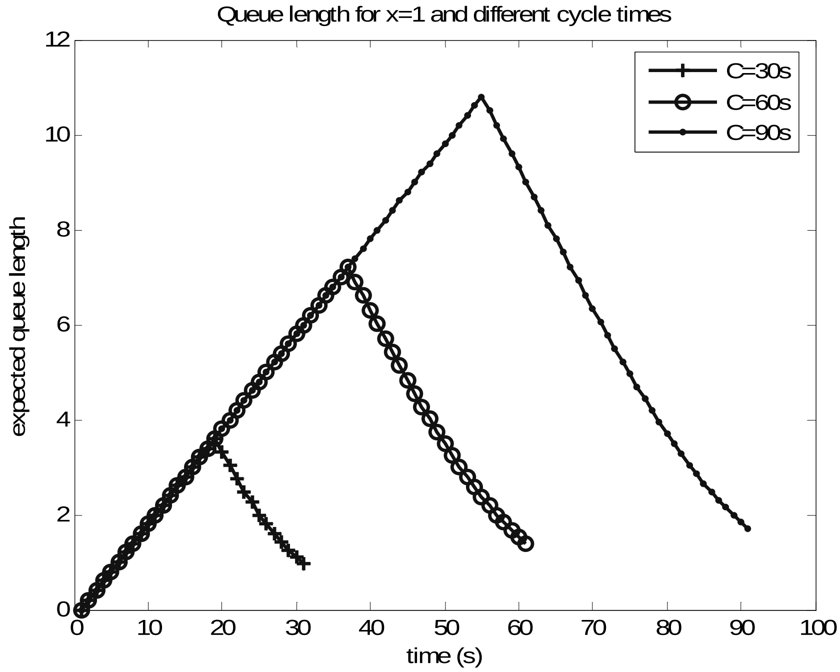

Figure 4.8: Expected queue at x=1

The compensation term does not have on the other hand particular role when the initial queue is large, as one can expect, since the probability of having zero queue at the end of the cycle is negligible (Figure 4.9). The expected value at the end of the cycle is very well represented by the deterministic term.

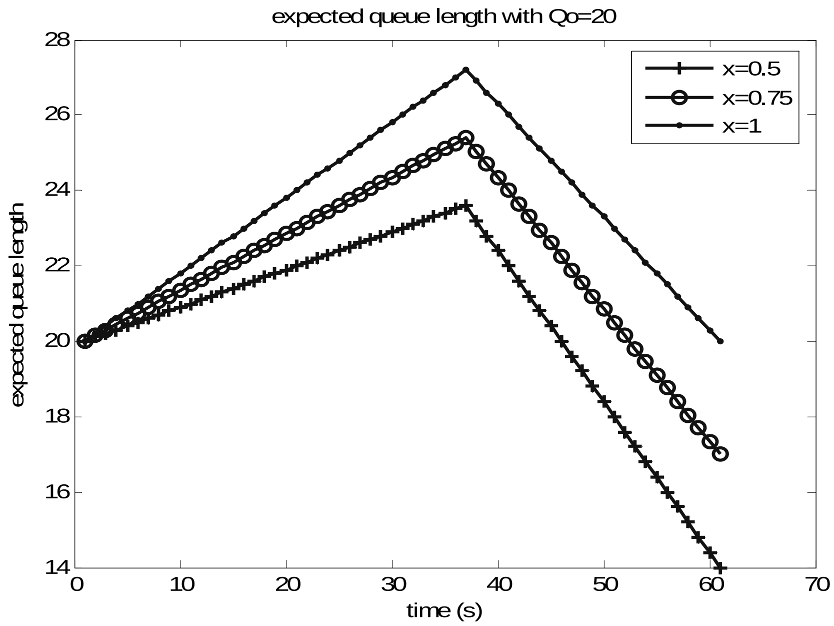

Figure 4.9: Expected queue for *Qo*=20

The new queue model formulation overcomes the gap between queuing models in conditions of undersaturated and oversaturated states, since no assumptions in this sense is made during the development of the model. Moreover, the model is valid for any assumed distribution of the arrival profile since the expression of the queue length is given without assuming any specific probability distribution function.

4.6 The control delay at one signal cycle

The inclusion of the compensation term in the expression of the queue length affects also the computation of the delay at one cycle. This allows one to compute the dynamic effects of the variability of the arrivals also in undersaturated conditions, and to give a smooth transition between undersaturated and oversaturated conditions also for the control delay. The delay formula proposed by Olszewski (Olszewski 1994) and presented in Section 4.4 is very useful if one computes from the probability distribution of the overflow queue the probability distribution of delays. The method summarized by Formula (4.12) considers the queuing process within a cycle as deterministic; this implies that the delay is simply considered with this method consisting of two components, the uniform delay component and the overflow delay component. This section shows that also the stochastic delay component should be calculated as incremental.

The expected total control delay is related to the dynamic of the queue length according to following relationship:

$$E[W] = \int_0^{t_C} E[Q(t)]dt \tag{4.25}$$

Where W is the total delay in a cycle t_C and $Q(t)$ is the queue length at time *t*. The queue length is in one single cycle a step function that increases with one at the arrival of a vehicle. If one takes the expectation value of the queue length it becomes a continuous function.

If one takes just the deterministic part of the queue he computes the uniform delay component:

$$E[W_1] = {(Q_0 + s.t_r)(Q_0 + q.t_r)} \big/ {2(s-q)} \tag{4.26}$$

Which reduces to the first term in Beckmann's delay formula if $Q_0 = 0$. The *random* delay component becomes:

$$E[W_2] = \int_{t_r}^{t_C} \sum_{l=0}^{[s(t-t_r)]} \Pr(q = l, t) \sum_{m=0}^{[s(t-t_r)]-l} (s \cdot (t - t_r) - m - l) \cdot \Pr(Q = m, 0) \cdot dt - \\ - \int_{(Q_0 + s \cdot t_r)/(s-q)}^{t_C} \{Q_0 + q \cdot t - s \cdot (t - t_r)/(s - q)\} \cdot dt \qquad (4.27)$$

This can be simplified to:

$$E\left[W_2\right] = \int_{0}^{t_C - t_r} \sum_{l=0}^{[s \cdot t']} \Pr(q = l, t' + t_r) \sum_{m=0}^{[s \cdot t']-l} (s \cdot t' - m - l) \cdot \Pr(Q = m, 0) \cdot dt' - 0.5(s - q) \cdot \tilde{t}^2 \qquad (4.28)$$

With $\tilde{t} = t_C - (Q_0 + s \cdot t_r)/(s - q)$ the undersaturated green time. Formulas (4.26) and (4.28) represent *exact* formulations of uniform and random delays within a cycle. Knowing the distribution of the initial queue length at the start of the red phase, the models presented in this chapter allow one to evaluate the dynamics of the queue and the delay during the cycle, explicitly including the effect of the variability of the arrivals in the dynamics of the queue length. This justifies the dynamics in the cycle-to-cycle process described in Section 4.3. In conclusion, a complete computation of the delay is now possible if the cycle-to-cycle Markov Chain model is combined with the newly developed model, as shown in the next section.

The Expression (4.25) is not simple to evaluate by hand and the relationship with easily obtainable traffic characteristics is absent. The degree of saturation is in fact not directly visible in the expression for the random delay. However, the calculation can very easily be executed by a rather simple computer program.

4.7 Comparison with other random delay formulas

The elaborate formulation of the queue and delay processes presented in this section limits its application opportunities. The advantage on the other hand is that it is ‘clean’, not influenced by any approximation. Two options are although possible: to use Formula (4.25) as a validation for existing formulas of random delay or to search for a simple formula that approximates this expression. This last approach will be followed in chapter 6 for the evolution of Q_0 , the initial queue at the start of a red phase. The dynamics of the expectation value and its standard deviations have been approximated very accurately by a mathematical function.

In this section we propose to follow the other approach, i.e. to use Formula (4.25) with the different random delay functions in the literature. For instance the well-known delay functions of Webster (Webster 1958) and McNeill (McNeill 1968) as representatives of the equilibrium delay models and Akcelik (Akcelik 1980) for the time-dependent models can be compared with the new expression for the delay.

If the probability distribution of the initial queue length Pr(Q=0) is known and the arrival is assumed distributed as e.g. Poisson, the random delay can easily be calculated numerically. Figure 4.10 shows that Expression (4.25) is consistent with Webster's delay formula for low degrees of saturation (x < 0.96).

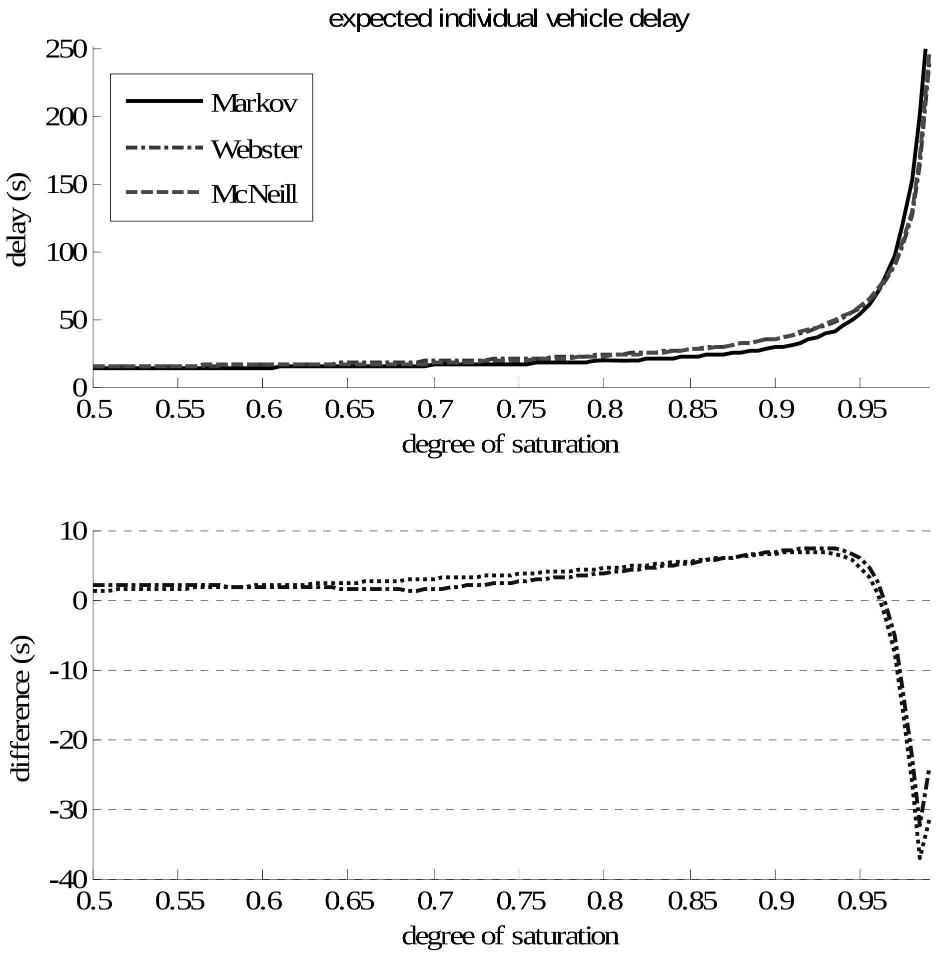

Figure 4.10: Comparison of the results from equation (4.25) with the Webster's and McNeill's formulas.

This error affects the estimation of delays at signals and consequently the correct design of signal controls. The newly developed models are supposed to overcome these limitations. To give an example of the effects of modeling exactly the effects of a variable demand in the optimization of a fixed control, Figure 4.11 shows the optimal cycle times (in terms of total delay minimization) computed with the well known

Webster's formula (Webster 1958) and the ones computed with the probabilistic approach.

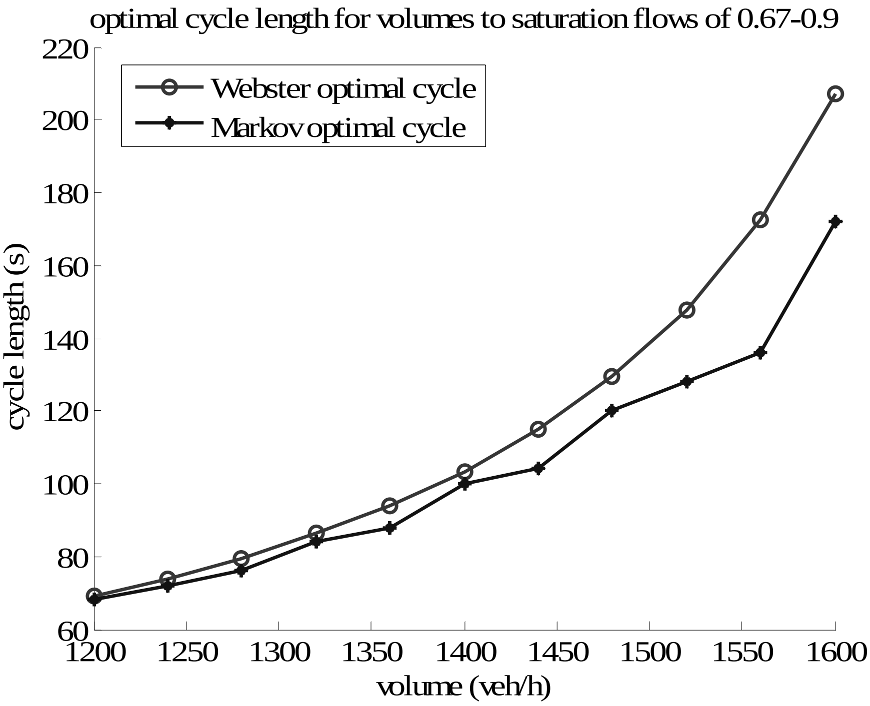

Figure 4.11: comparison between Webster and Markov optimal cycle lengths

The optimal cycle time method proposed by Webster (Webster 1958) in undersaturated conditions is derived with the assumption of infinite stationary demand conditions. This implies that overflow queues in the Webster formula are considered reaching instantaneously the equilibrium value. In practice, for degrees of saturation approximately near 1 one should need very long periods of stationary demand conditions to reach these equilibrium values, which is not a realistic assumption. In the Webster's method the expected delays tend to infinite at conditions of traffic near the capacity. In reality stationary demand conditions may exist only for a limited time period and queues and delays are, within this period, finite and – in case of a degree of saturation above 0.8 - considerably lower than the ones estimated with the Webster's delay function. Optimal signal settings are therefore different if one replaces the theoretical expression of delays with a time-dependent model. The Webster's method in fact optimizes green and cycle times in cases of low volume to saturation flow ratios consistently with the results of the Markov model. However, it overestimates the cycle length when the demand is near the capacity.

The reason of this overestimation error can be found in the second term of the delay equation proposed by Webster and also introduced in chapter 3. This method computes the expected stochastic delay under the assumptions of no initial queue and uniform arrivals. One drawback of this equation is that it cannot be used in conditions of flows larger or equal to the capacity, since the second term gives an infinite delay for $x = 1$ and it cannot be applied if $x > 1$.

The difference between Markov Chain and Webster's models appears in the expression of the optimal cycle length based on delay minimization (Webster's formula has been given in Chapter 3). Figure 4.12 shows the expectation value and the standard deviation of the control delay using different cycle lengths. The right hand picture displays the standard deviation (continuous lines) and the corresponding average values (dotted).

The knowledge of the variability of delays can be also a helpful tool also in the optimization of signals. Figure 4.12 shows the behavior of the standard deviation, which decreases with increasing the cycle length. Comparing the behavior of the standard deviation with the corresponding expectation value one can estimate the coefficient of variation (ratio between average and standard deviation) having a measure of the reliability of delays given the total expected demand of the signal in time.

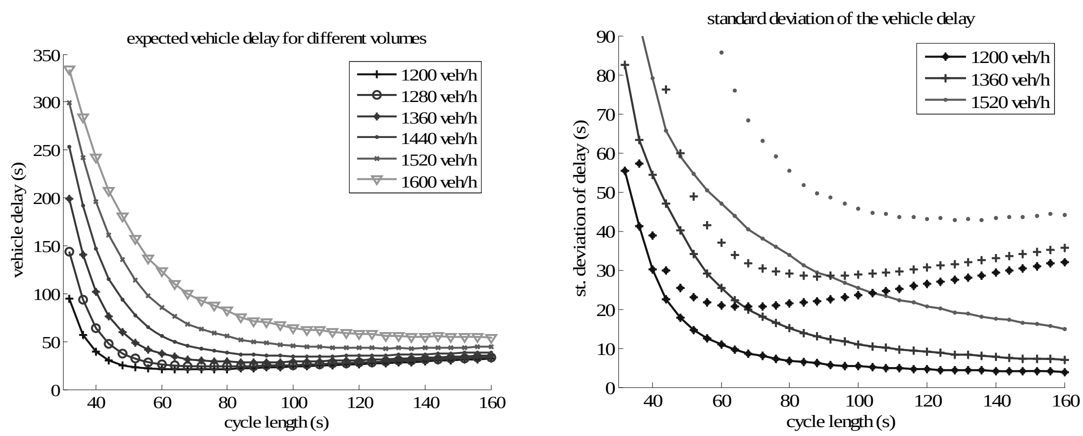

Figure 4.12: Expectation value and standard deviation of the delay with different cycle times

For low demand conditions the standard deviation of the delay at the optimal cycle length is small. For higher demand conditions, if one increases the cycle length with respect to the optimal value, the expectation of delay does not increase consistently, while the standard deviation is strongly reduced. In conclusion a control policy with particular regard to the reliability of a traffic control should design, in nearly congested conditions, longer cycle lengths than the computed optimal ones.

According to the conclusions drawn in this section the optimal cycle length computed with the Markov method are smaller than the ones computed with the Webster method. This is especially evident in the busiest period, where the optimal cycle length is considerably lower than the Webster's optimal length. Average delays within each period are also lower than the ones computed with the Webster optimal cycle, and in the busiest period they are even less than half.

Table 4.1 – Assigned demand and resultant average delay with Markov chain and Webster methods

Flow (veh/h)	**600**	**700**	**750**	**800**	**750**	**700**	**600**	**500**
Webster opt. cycle	60	90	120	150	120	90	60	45
Webster delay (s)	22.4	34.6	46.8	73.1	46.8	34.6	22.4	15.2
Markov opt. cycle	58	90	126	126	126	86	58	50
Markov opt. delay	19	30.1	31.1	31.8	30.9	30.2	19.5	14.1

Finally comparison between the time-dependent formula of Akcelik (Akcelik 1980) and the Markov model is done in Figure 4.13. It seems that the heuristic approach made by Akcelik gives very consistent results with respect to the Markov model.

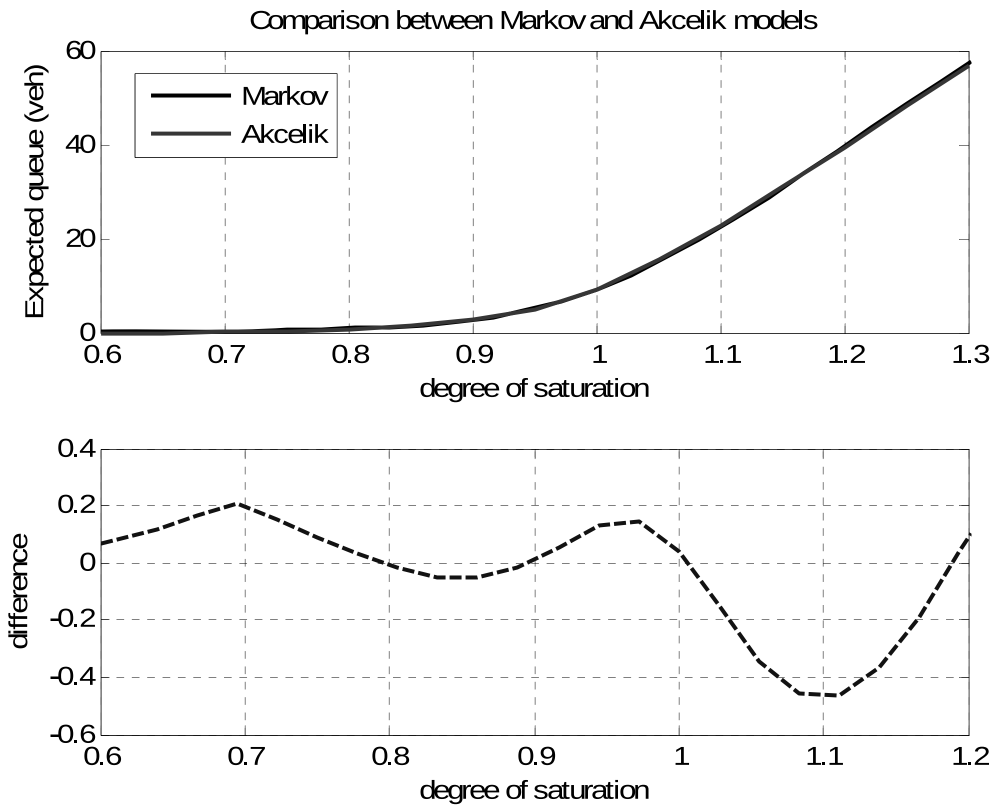

Figure 4.13: The comparison of equation (4.25) with Akcelik's model

4.8 Conclusions

Although fixed time controllers are considered as old-fashioned while most intersections are controlled by closed loop controllers, the traffic dependent control becomes nearly fixed time during peak hours. That makes a good model for delays of fixed time controlled intersections still important. This chapter provided a methodology to simulate the dynamic and stochastic behavior of queue lengths and delays at pre-timed traffic control signals based on Markov renewal process theory. This method allows the analyst to estimate and predict the dynamic evolution of queues and the propagation of their distribution in time, quantifying the uncertainty around this estimation and/or prediction.

Delay at fixed-time controlled intersections has been a study subject for many years already. Several mathematical expressions have been derived to represent the so-called random delay component, the delay caused by the stochastic character of arriving traffic. The new formulation of this random delay component in this thesis is derived without any assumptions about the statistical properties of the arrival process, apart from the assumption that the arrival distribution is uniform over the whole cycle. It fits for Poisson, binomial, Normal distributed arrivals. The numerical value of the new random delay component can easily be computed and compared with approximated expressions of other authors.

More important than the possibility to calculate the random delay with a more general model than the existing ones is the insight that the derivation gives in the process that causes the random delay. The assumption used by some authors that the queue should be represented by a step function appears to be superfluous. The stepwise character of the delay is transformed to a smooth character of the expected delay, linearly increasing in the red-phase and the first part of the green phase. The expectation value of the queue in the green phase shows a non linear character as soon as the tail of the probability distribution comes close to zero. This phenomenon causes the overflow delay.

Numerical evaluation of the delay model presented in this thesis allows one to evaluate the dynamic character of queues with the variability of the arrival process. This feature has been described so far by computing the expected overflow queue at the end of the green phase and by using a mass-balance equation. The approach used in this chapter is to compute this value by catching the dynamics from any point in time. This information is very important for example to consider spillback and gridlock effects.

Comparison of the model with previously released analytic formulas shows that the latter are particularly incorrect at conditions near capacity and that they are not suited for analyzing the time-dependency of delays in e.g. peak hour analyses. A new time-dependent analytic formulation is therefore needed to solve this issue. Chapter 6 is dedicated to the development of such model.

5

Dynamic and stochastic aspects of queues and delays at signals

5.1 Introduction

The variability of demand and supply systems to the behavior of queues and delays has been shown in Chapter 2 to have large impact especially in conditions near capacity. Travel time data analysis has shown that the standard deviation of travel times is as large as its mean value at congestion. This variability transmits large uncertainty to the queue distributions and consequently to the random component of delays. Expectation values are therefore determined by a wide distribution.

All models for the expectation value of queues were found to have a deficiency in dealing with the dynamic and stochastic character of overflow queues in conditions near capacity. Chapter 4 has solved this by providing an exact time-dependent formulation of the expected queue length and the control delay within a signal cycle. This formulation adds new insight into the dynamics of the incremental component of the delay with respect to models based on heuristics (e.g. (Akcelik 1980)).

Although the uncertainty in the queue length and delay estimation may be as important as their expectation values, little interest has been found in literature on this issue. Among the studies presented in Chapter 3, only very few discuss the variability of queues and delays at signals and only one model has been developed (Fu 2000).

This formulation describes the queuing process as a Markov Chain (Markov 1971), i.e. the queue length at one time step depends only on its distribution at the previous time step. This approach has been used other times in the past to compute the queuing process as a cycle-to-cycle process (e.g. (Van Zuylen 1985), (Olszewski 1990)) and to analyze the expected value of the queue length as a dynamic process.

Less importance has been given to this methodology to analyze also the behavior of the variability of queues. In fact, this method enables one to calculate the statistical distribution of queues under controlled conditions in a few seconds of computation. The adopted method does not require the specification of each vehicle characteristic in the same detail of microscopic models, but it requires only the specification of probability distributions of the input variables to determine the queuing behavior in time. One may question the validity of using this methodology based on rigorous mathematical relationships to evaluate this behavior. Unfortunately, it is difficult to find a valid dataset, which can be used to solve this issue. Only comparison with microsimulation seems possible, as discussed in Chapter 7.

The scope of this chapter is to analyze the statistical properties of queues and delays by using the Markov Chain process. The chapter is structured as follows. Section 5.2 describes the reasons for adopting a probabilistic approach to operate the analysis of the dynamic and the stochastic aspects of queues. In Section 5.3 the probabilistic method described in Chapter 4 is applied to the fixed time, isolated signalized intersection problem under various conditions of traffic and initial states. Analysis of the variability of queues is made in Section 5.4 while in Section 5.5 the analysis of the queue behavior with non-uniform arrivals is presented. Section 5.6 finally gives a synthesis of this chapter.

5.2 Reasons for a probabilistic approach

Chapter 3 has shown that available time-dependent models have been developed under quite limiting assumptions. The simulation of a peak hour requires for example a model, which enables one to compute not only expected delays and queue lengths but also the duration of such queues. This would improve the evaluation and optimization of traffic performances in traffic control or information problems. Furthermore, the knowledge of the distribution of travel times is required in several contexts, like in information problems, or in reliability and robustness analyses. Therefore, a model for the standard deviation and a characterization of queue length distributions is needed, contributing to the evaluation and prediction of travel time uncertainty. In conclusion the research questions, which motivate the approach described in this chapter, are the following:

1) To formulate a stochastic time-dependent model that describes the dynamic evolution of expected overflow queues with variable conditions of traffic;
2) To provide a stochastic time-dependent model also for the evolution of standard deviation of queues;
3) To analyze the evolution of the queue distributions in time and to characterize such distributions with a known probability density function.

5.2.1 Mesoscopic models

To analyze the statistical properties of queues and delays at signal one needs to have a large dataset, both because different conditions of traffic need to be analyzed, i.e. different demand conditions, control settings, initial queues etc. and because the number of repetitions required to have an acceptable estimate of average and standard deviation of the queue length increases with the variance-to-mean ratio. The development of a time-dependent model of the expectation value of queues requires several observations under the same prevailing conditions. Observations with these assumptions are very difficult and expensive to gather. The number of observations required is unrealistic both for time and monetary reasons, and because the arrival distribution is not controllable. Also measurement errors should be taken into account, since they also provide input for a travel time model. (Teply 1989) studied two approaches concluding that delay cannot be precisely measured and that a perfect match between the results of an analytical delay formula and delay values measured in the field cannot be expected. Thus, even if one has budget and time to collect data from any data collection system, there is still uncertainty on the correctness of this dataset. Alternatively, one may use microscopic simulation programs to generate these observations artificially. The opportunity to deal with the characteristics of any single vehicle increases the flexibility and power of microscopic simulation programs, but at the cost of an increased complexity, the need of extensive calibration, the risk of over-fitting and a reduction of computing speed.

Both real observations and microsimulations are difficult, time consuming and very expensive to be collected. Under some conditions, the standard deviation of the queue can be of the same magnitude of the average value. Therefore, to obtain an estimate for example with 10% accuracy in these conditions, hundreds of observations are required. (Troutbeck 2000) showed that, while analyzing the validity of the models of (Newell 1971) and Kimber and Hollis (Kimber 1979), even with 500 simulation runs of a microscopic model, the resulting curves representing the average conditions are not smooth.

Alternative to microsimulation is represented by the class of *mesoscopic models*. The Federal Highway Administration (FHWA 2006) defines mesoscopic models as follows:

> Mesoscopic simulation models combine the properties of both microscopic and macroscopic simulation models. Mesoscopic model travel simulation takes place on an aggregate level and does not consider dynamic speed/volume relationships. As such, mesoscopic models provide less fidelity than the micro-simulation tools, but are superior to the typical planning analysis techniques.

Mesoscopic models use as input probability distributions, computing the probability distribution of some traffic characteristics applying some mathematical conditions that relate these characteristics (see also Appendix A).

5.3 Evolution of queue length distribution in time

Firstly, the assumption of stationary demand conditions for the whole evaluation period and the hypothesis of no queue at the start of the evaluation period and deterministic capacity are assumed. Hypotheses of constant departures and zero initial queues are later stressed in this section.

5.3.1 Influence of stochastic volume-to-capacity ratio

Figure 5.1 shows the average length of a queue for different volume-to-capacity ratios computed with the Markov method described in Chapter 4.

As mentioned in Chapter 4, for low volume-to-capacity ratios ($x \leq 0.7$) the chance to observe a residual queue at the end of the green phase because of random arrival rates is quite small. Consequently, delays are primarily given by their uniform component. The higher the demand rate, the higher the chance a residual queue due to its random nature is observed at the end of the green phase. Volume-to-capacity ratios in between [0.8,1.2] are influenced consistently by this fluctuation of volumes (as one can notice in the above figure) and the random delay represents the main component. For v/c larger than this range the queue increases linearly and the stochastic component is again negligible.

Figure 5.2 compares the Markov model with Akcelik’s model for a volume-to-capacity ratio of 0.95. The models give similar results although from visual inspection the first tends to give slightly higher results than the latter. If several cycles are investigated, this difference reduces until both models reach equilibrium in undersaturated conditions, which is in accordance with the steady-state expressions of McNeil (McNeill 1968) and Miller (Miller 1968).

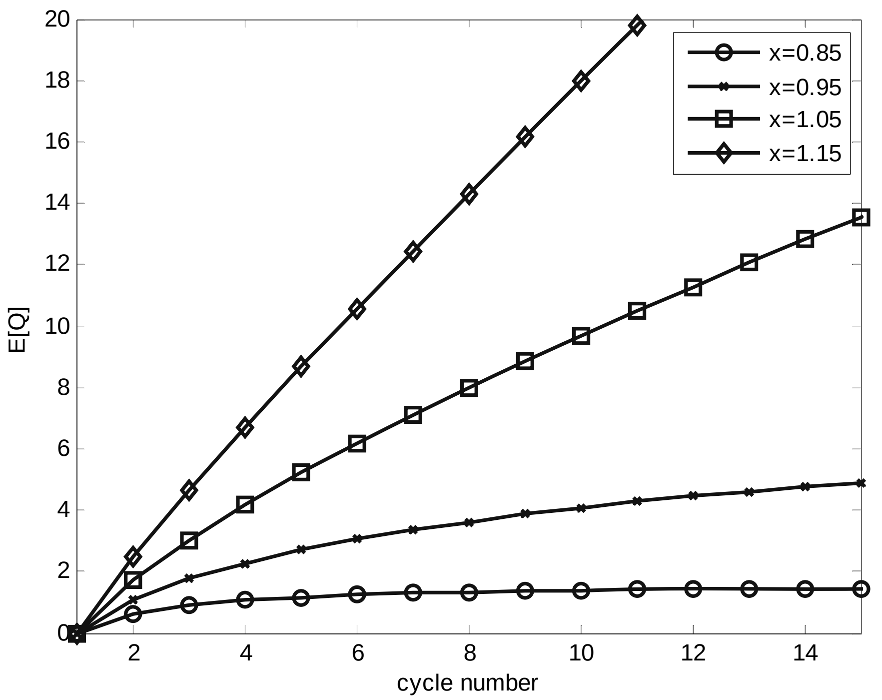

Figure 5.1– Average overflow queue length for different degrees of saturation

Empirical models based on the static ones have been developed in the past (e.g. (Kimber 1979)) by applying the coordinate transformation technique. Akcelik (Akcelik 1980) developed a model with the property of expressing both undersaturated and oversaturated conditions with a single formulation, as seen in Chapter 3. This formula is the most used model in case of flows approaching capacity and initial queue $Q_0 = 0$. On the other side it cannot reproduce queues increasing or decreasing from a non-zero initial value. A recent extension of the model for including the effect of an initial queue can be found in the aaSIDRA manual (Akcelik 2002). The authors add to the flow rate q a term that takes the residual queue generated in the previous intervals into account. The model uniformly distributes the extra flow over the calculation period adding the ration Q_0 / T, where Q_0 is the expected residual queue from the previous time step and T is the time step length. Akcelik's model computes the expected queue length at the end of the evaluation period, set to 15 minutes. Capacity is assumed deterministic in this model.

Comparison between the analytical formulae adopted by the Australian, American and Canadian manuals and with other simulation approaches (shock wave theory, microscopic models) has already demonstrated the consistency of such models under all degrees of saturation if the above assumptions are made. Dion et al. ((Dion 2005) Figure 10, page 17) reports the comparison of these models in comparison to the results of a microscopic simulation program, INTEGRATION (Van Aerde 2001). All Manuals use time-dependent models that are structurally similar to the Akcelik's function.

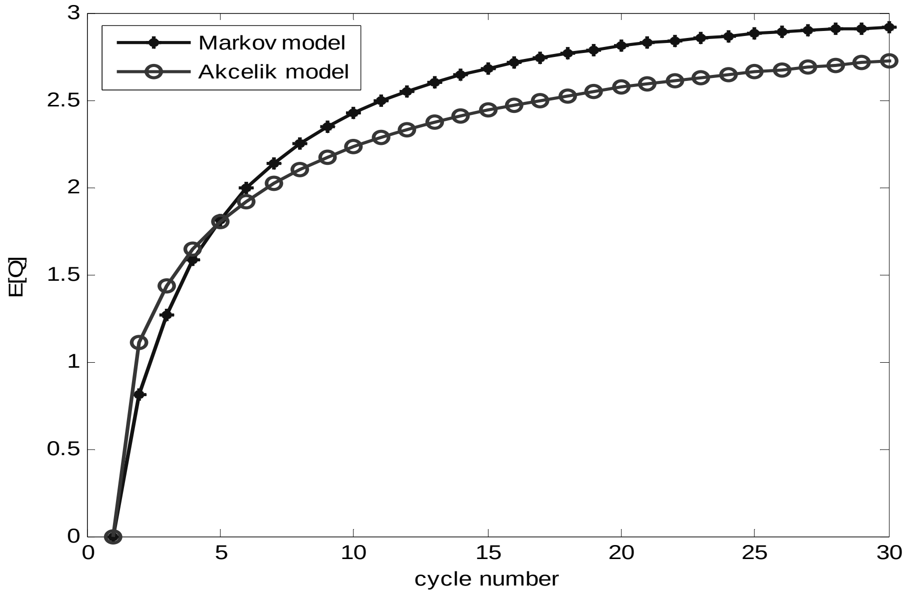

Figure 5.2– Markov and Akcelik overflow queue models for an under saturated condition

In oversaturated conditions the Markov model provides average overflow queues that well represent the linear deterministic behavior. Figure 5.3 shows this comparison in case of x=1.15. The stochastic behavior influences the expected conditions only for a small part during the first cycles.

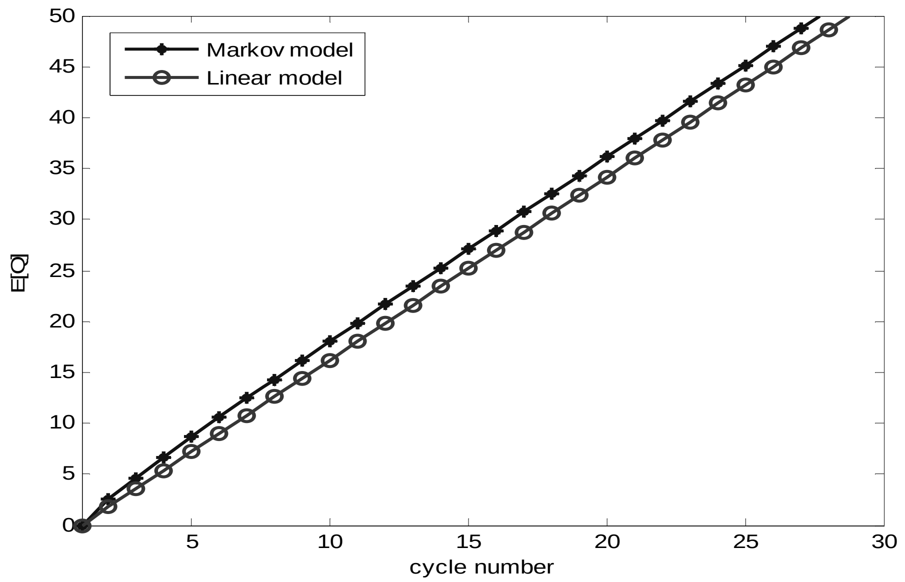

Figure 5.3 – Markov and linear overflow queue models for an oversaturated condition

This difference in behavior can be explained by looking at the probability distribution Figures 5.4-5.5 display the evolution of probability distributions in time for respectively x=0.90 and x=1.15.

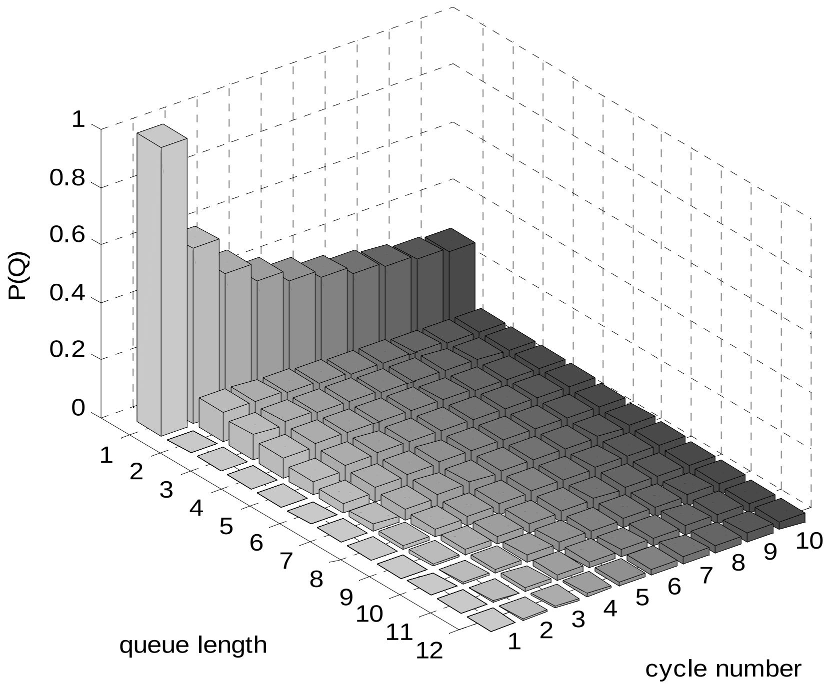

Figure 5.4 – Evolution of probability distributions for *x*=0.90 and Q_0=0

For under saturated conditions the highest chance is attributed to the zero state, which begins with the assumed $P(0) = 1$ and gradually decreases until it reaches, in the example shown, a value of around 50%. The distribution seems to be well approximated by an exponential distribution. Although the chance for the queue to be zero is the predominant state, in nearly half of the chances the queue will not be zero. The curvature of the average state is then determined by this gradual transmission of probability from the zero to the non-zero states.

The oversaturated case shows a different behavior. The probability for a queue being zero becomes already unlikely from 5 cycles (0.001%). The distribution is well approximated by a Normal distribution, in accordance with the conclusions of Newell (Newell 1971), who characterized the queue length distribution as Normal using the statistical law of the weak numbers. The expected value of the queue moves cycle by cycle away from the zero state. This explains the linear behavior showed plotting the average value. Accordingly, the distribution in under saturated cases can be approximated by a Normal distribution too, bounded to be non-negative (Viti 2004).

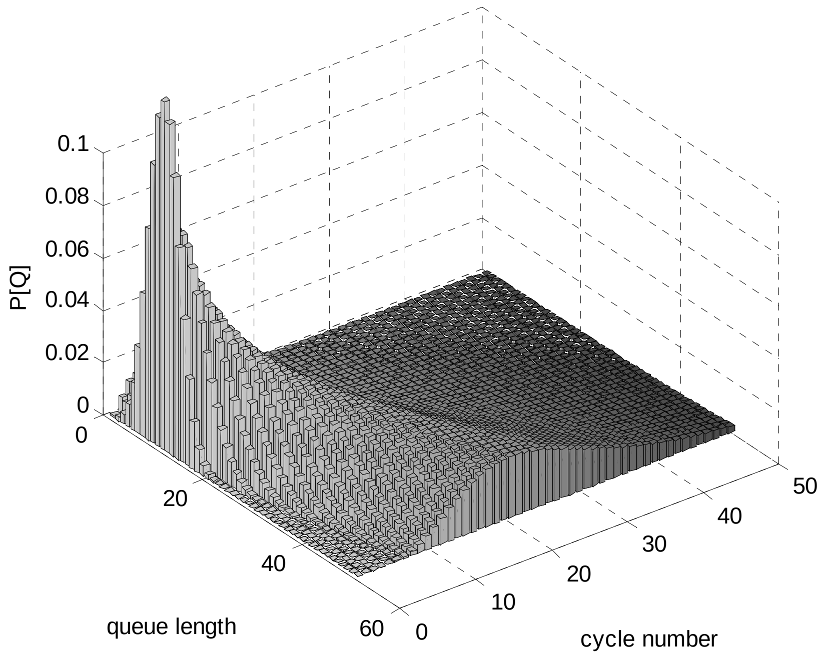

Figure 5.5 – Evolution of probability distributions for x=1.15 and Q_0=10

5.3.2 *Influence of stochastic departures*

The number of vehicles departing at each cycle depends primarily on the effective green time assigned to each arm. Variability of vehicle headways and reaction times, the presence of heavy vehicles, the occurrence of incidents, or simply distraction or slow reaction of travelers may also influence consistently the number of vehicles departing at each cycle and consequently the capacity.

Olszewski (Olszewski 1990) concluded from some observations in the city of Singapore that the number of departures within a cycle can be characterized by a Binomial distribution with coefficient of variations within the range $[0.03, 0.08]$ and suggested the computation of stochastic capacity using this distribution and assigning a coefficient of variation of 0.10 to include somewhat a safety margin. Therefore, the parameters p and n of the Binomial distribution are computed by fixing this coefficient of variation (CoV) and the average departures d with the following system of equations:

$$\begin{cases} np = d \\ \sqrt{\dfrac{(1-p)}{np}} = CoV \end{cases} \tag{5.1}$$

The first condition is derived from the definition of average of a Binomial. For example for average departures of 12 veh/c and $CoV = 0.10$ one obtains the values $p = 0.88$ and $n = 14$. Under these assumptions, the average random queue for degrees of saturation in between $[0.80, 0.98]$ increases with a maximum increase of 18-20% of the corresponding deterministic case. This computation is higher than the computation made by Olszewski (Olszewski 1990), which estimated an error of maximum 7%.

Figure 5.6 compares the average queue length computed with deterministic and with stochastic capacity for a degree of saturation of x=0.97 and for 1 hour simulation.

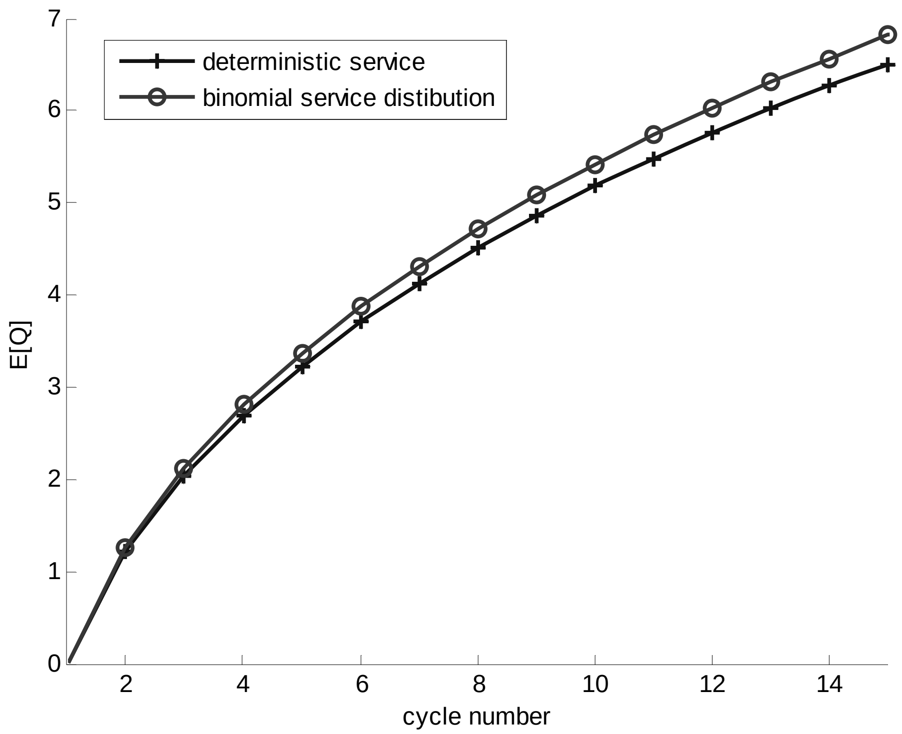

Figure 5.6 – Behavior of average queue length with deterministic and stochastic capacity

The larger the degree of saturation the smaller the effect of stochastic capacity influences the average queue until under oversaturated conditions the influence of stochastic capacity becomes negligible as much as the stochastic arrivals.

5.3.3 Influence of an initial queue

So far, the queue length distribution has been analyzed starting from the condition that no queue exists at the starting of the evaluation period. Under this condition, the overflow queue increases with time for every condition of traffic, as seen in Section 4.7. If an initial queue is assumed, the overflow queue has a different behavior, decreasing when the initial value is larger than the equilibrium value computed in the zero initial queue case. The presence of this residual queue causes an extra delay to the vehicles approaching the intersection for several cycles onwards, as demonstrated in chapter 4.

Catling (Catling 1977) suggested computing the evolution of the decreasing queues with the simple linear function described by Formula (4.1) or with the following expression if steady-state equilibrium is computed:

$$Q_t = \max\{Q_0 + (x-1)\cdot ct, Q_e\} \quad (5.2)$$

where c represents the average number of departures per cycle and Q_e is the equilibrium value computed under steady state conditions. This function has a descending trend as long as Q_t is larger then the equilibrium value and later stabilizes at the equilibrium value. The Highway Capacity Manual (TRB 2000) specifies a third component in the delay model, which is caused by this initial value. The component of the total delay caused by the residual queue assumes that the queue decreases linearly following Formula (5.2).

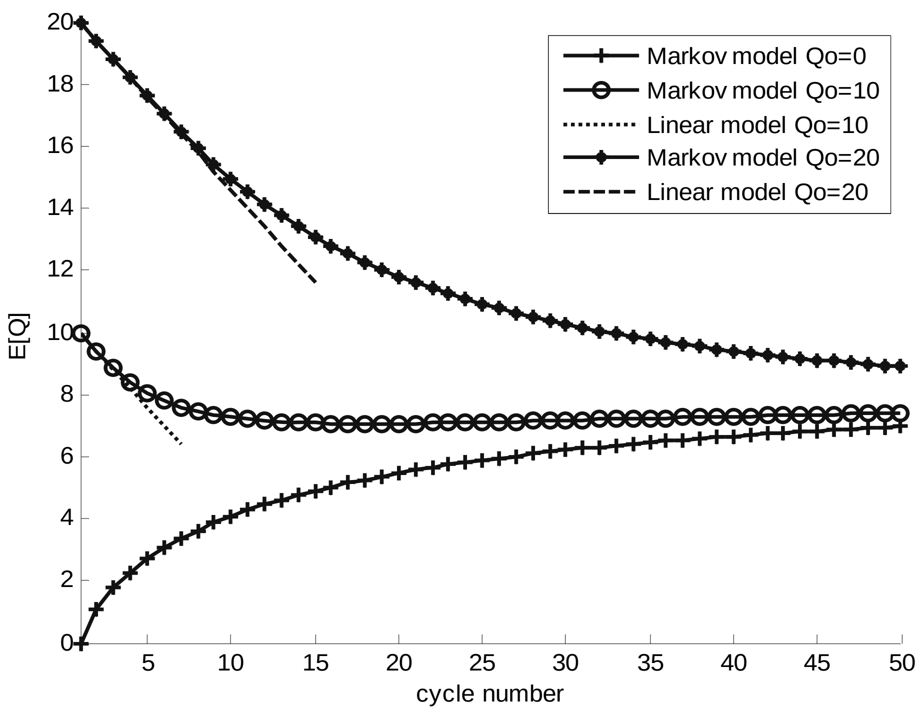

Figure 5.7 – Behavior of average queue length with different initial values and x=0.95

The decreasing behavior is initially computed with the Markov model assuming deterministic initial values. This assumption can still have a practical use; if the traffic manager has somewhat the information about the actual queue length at a certain time T (for example using cameras) he/she can then estimate how likely the average queue evolves from that time on, and give this information to the travelers. Therefore this method can be helpful for online short-term predictions. Figure 5.7 shows the evolution of the queue length computed with the Markov model for a degree of saturation of 0.95, effective green time of 24 seconds, cycle time of 60 seconds, total evaluation period of 100 cycles and with a deterministic initial queue of respectively 0, 10 and 20 cycles.

The queue evolves initially according to the deterministic function when the initial queue is larger than the equilibrium value while it keeps on increasing linearly if it is larger or equal than 1 and exponentially if the initial value is smaller than the equilibrium value.

The computation of an initial queue that is close to the equilibrium value (the case of 10 vehicles) gives a strange result of in terms of expected value of queues, which first decreases reaching a minimum below the equilibrium value and after some cycles increases again till equilibrium. The behavior is explained by considering the evolution of the distribution and noticing that the probability of a queue being zero is initially zero and the system needs some cycles before this probability could assume the same importance as in the case of queues that start with zero initial values. Since the probability of overflow queues for these first cycles are well approximated by a Normal distribution, the average of queue lengths for these initial cycles follows the linear evolution computed by Formula (5.2).

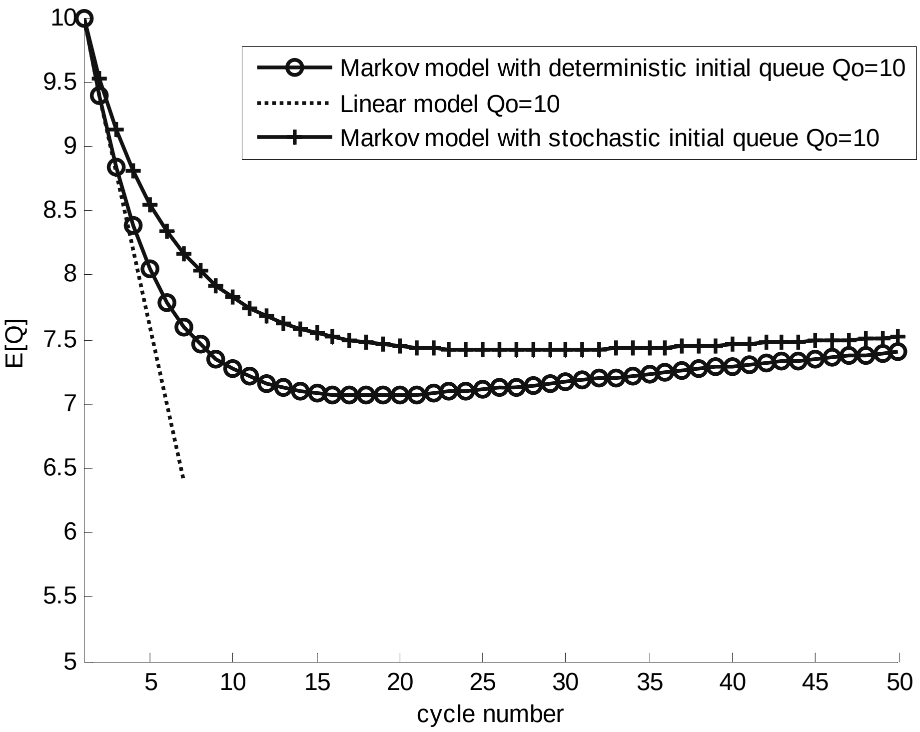

Figure 5.8 – Behavior of average queue length with deterministic and stochastic initial value

Starting from any initial value, the queue reaches a common equilibrium and the distribution of overflow queues is the same, thus the steady-state condition is an *absorbing state* to the system (see also Appendix A). Moreover, this finding implies that the initial queue value does not influence the final value at equilibrium. Looking at the evolution starting from 20 vehicles, the queue follows the deterministic behavior for about 10 cycles and reduces less sharply afterwards, following an asymptotic curvature till equilibrium. This implies that, if the queue behavior is simply computed by Formula (5.2) this will lead to an underestimation of the total initial queue delay.

The error increases if the initial queue is stochastic and described by a probability distribution. Figure 5.8 shows the different behavior computed with the Markov model if a standard deviation of the initial queue is assumed equal to the average. Initial distribution is assumed Normal, bounded to be non-negative. Until the standard deviation is small with respect to the mean, there is very little chance that one can observe queue at the end of the green phase and the queue follows the linear deterministic behavior. If the standard deviation of the initial value is comparable with the mean the probability of a zero state is already large enough from the first cycles to influence the evolution of the expectation values.

5.4 The variability of the overflow queue length

The influence of the standard deviation on the expected value of the queue makes an analysis of the same standard deviation necessary and creates the need for an analytical model for it as much as it is needed for the average value. The analysis of the standard deviation follows the same criterion used in Section 5.3, starting with the simplest case of stationary demand, deterministic capacity and zero initial queue.

5.4.1 Evolution of the standard deviation

The evolution of the standard deviation in time does not differ much from the average in under saturated conditions. Given the absorbing property of equilibrium, the standard deviation tends to reach also equilibrium (as shown in Figure 5.9 for the case v/c= 0.95).

For conditions of traffic near the capacity the standard deviation of the queue is always slightly higher than the average. Therefore, the uncertainty around the queue evolution in these conditions is very high. Since such conditions of traffic are often met in practice, the computation of the variability of queues can be useful, especially to compute the chance to observe spillbacks.

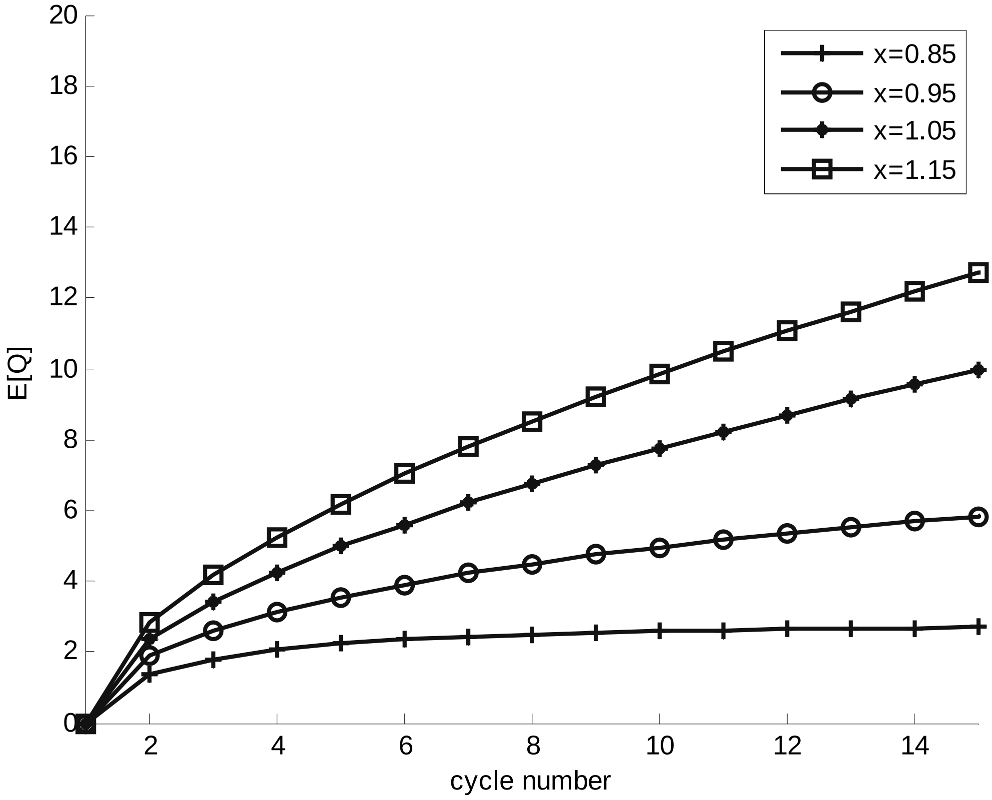

Figure 5.9 – Behavior of standard deviation of the overflow queue length

The difference between average and standard deviation decreases with increasing volume-to-capacity ratio. As already pointed out in Van Zuylen and Viti (Van Zuylen 2003), the relationship between equilibrium value of the average and the standard deviation can be approximated by adding a simple multiplicative factor, $\varsigma(x)$ (as displayed in Figure 5.10):

$$\sigma[Q_e] = E[Q_e] \cdot \varsigma(x) \tag{5.3}$$

The expression of the multiplicative factor $\varsigma(x)$ is the following:

$$\varsigma(x) = \left(x + \frac{1-x}{0.15} \right) \tag{5.4}$$

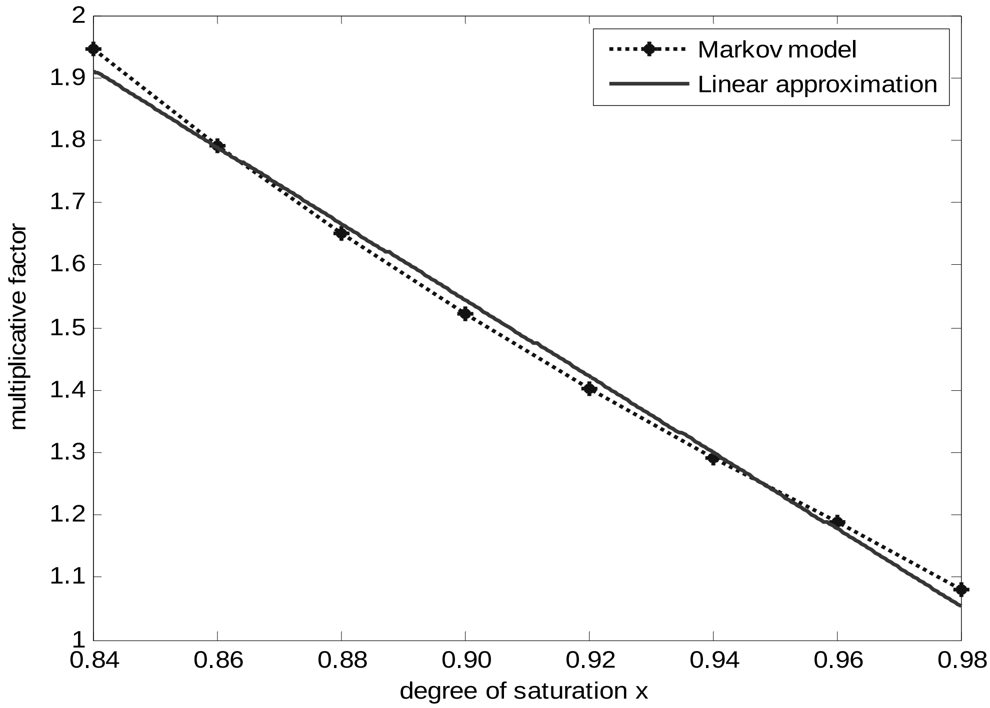

Figure 5.10 – Relationship between the multiplicative factor ζ and the volume-to-capacity ratio

Looking at the behavior of the standard deviation in oversaturated conditions (as displayed in Figure 5.9 for the case of v/c=1.05 and 1.15) the value becomes gradually smaller in time with respect to the corresponding average. Moreover, the evolution assumes a quadratic form. This conclusion is straightforward if one takes into account the relationship between average and standard deviation, and that the computed average is linear as seen in Chapter 4. For a zero initial queue and Poisson arrivals the standard deviation in undersaturated conditions is expressed by the following formula:

$$\sigma\{Q_t\} = \sqrt{xct} = \sqrt{at} \qquad (5.5)$$

Under the assumption of a positive residual queue the standard deviation in overflow conditions does not change in shape, since the chance of observing a zero queue at the end of the green phase is even smaller than the zero case. Formula (5.5) is thus still applicable.

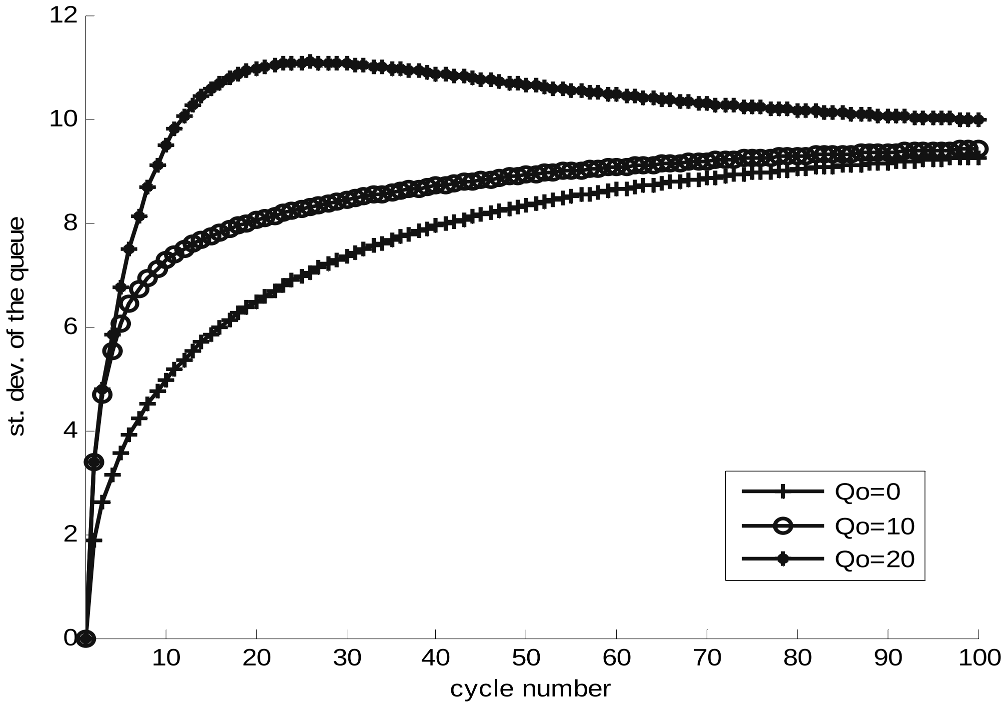

Figure 5.11 – Evolution of the standard deviation for different deterministic initial queues for x=0.95

In undersaturated conditions the initial queue influences the shape of the standard deviation, as one can see from Figure 5.11. In fact, if the initial value is assumed deterministic the distribution needs some cycles before a zero queue becomes the most likely state. Initially the standard deviation increases according again to Formula (5.5). As far as the chance to observe a zero overflow queue increases, the standard deviation decreases and asymptotically tends to the same equilibrium met with zero initial queues. This can be explained by looking at Figure 5.12. During the first cycles the variability of the queue distribution increases but the chance to observe a zero queue is still very small. After 30 cycles the distribution is very flat, while as long as the distribution evolves in time the probability of observing no overflow queues increases and it becomes the largest value. The standard deviation would not have this decreasing behavior if queues were not bound to be non-negative.

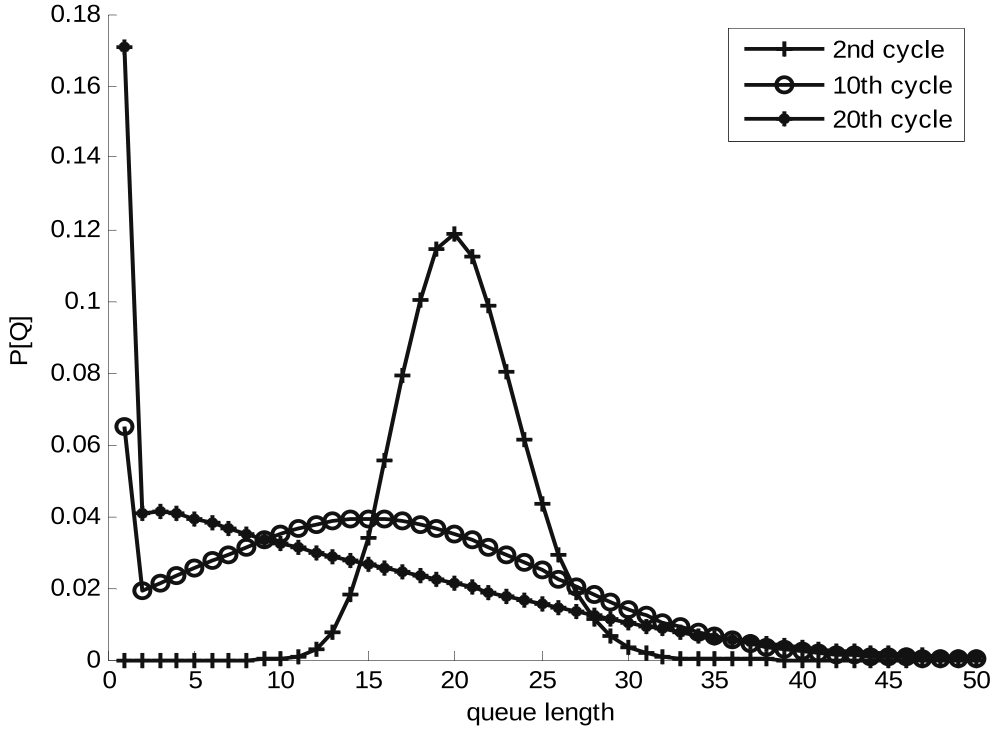

Figure 5.12 – Evolution of the overflow queue distribution for x=0.95 and Qo=40

A practical application of these results can be in the estimation of the probability of spillback to occur. The probability distribution at the 150th cycle in the figure is statistically in equilibrium. If a maximum number of passenger car units is assumed upstream the road section, N, one can compute the probability of spillback by computing the cumulative distribution value of N. For example, a section that allows 30 vehicles to queue up has from the example in figure 5.12 a 2% chance to be fully occupied, while a section of 20 vehicles has more than 10% chance.

Inversely, in the design of a road section one may fix a maximum chance of spillback to occur (e.g. 1%), and derive from the above probability distribution the minimum length according to this value. A similar method has been recently developed in the design of weaving sections (Ngoduy 2006). The method applies in the same way also to signalized road sections.

Third opportunity given by computing the probability distribution in time of overflow queues is the possibility to calculate the chance of spillback in time, which is particularly appealing in peak hour analyses. In these analyses one may be interested for example in evaluating the time of the whole peak period in which the probability of spillback is above the maximum allowed.

5.4.2 Relationship between expected value and variance

The interdependency between expected value and variance of the overflow queue is particularly evident if one considers the results presented so far in this chapter. If the standard deviation passes from zero to a finite value, the behavior of queues has gradually an exponential behavior instead of the linear deterministic one. On the other hand, if the initial queue changes, also the behavior of the standard deviation clearly changes. It could be useful to determine then the proportion of the variance-to-mean above which the exponential behavior becomes predominant. With this aim, a simulation of two consecutive periods is made, in which an undersaturated condition of traffic follows a condition of traffic determined by a higher degree of saturation. This degree of saturation can be still under saturated and oversaturated.

Three different initial demand conditions are simulated in Figure 5.13, respectively corresponding to a degree of saturation of 0.95, 1.05 and 1.15 for the first 30 cycles, while the second period, assumed equally long, is fixed to a degree of saturation of 0.90. Control settings do not change in between one period and another.

One may observe that the smaller the variance-to-mean ratio at the end of the first period of constant average arrival rate, the longer the initial linear behavior describes the overflow queue evolution in time at the starting of the second period of constant average demand. In conclusion the coefficient of variance influences the behavior of the queue for the subsequent period. If the average is sufficiently larger than the equilibrium value and the standard deviation is less than 30% the average the linear behavior is certainly observed for the first periods. If instead one of these conditions does not hold, the queue will more likely start from the first cycles to show an exponential evolution.

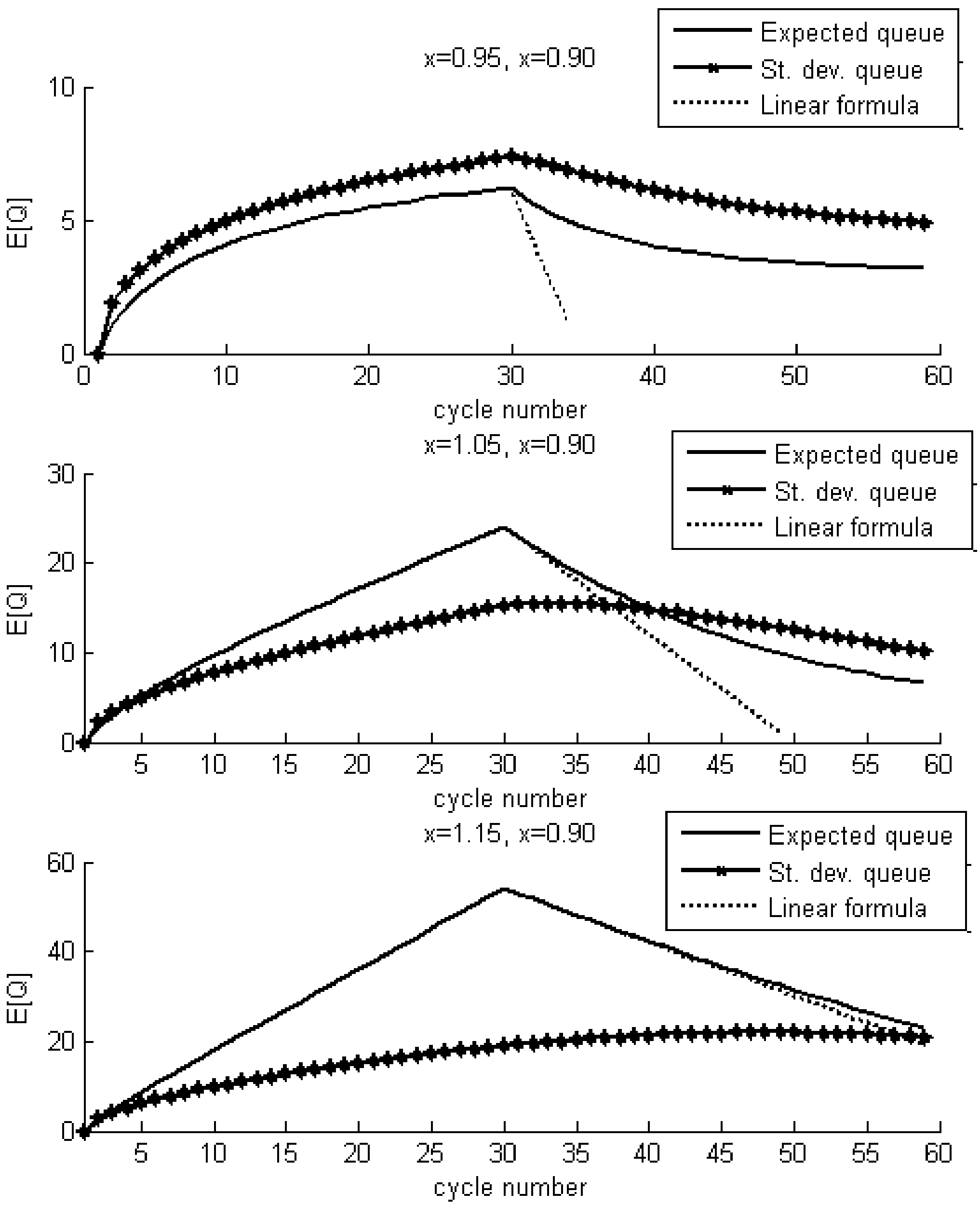

Figure 5.13 – Evolution of mean and standard deviation with two different periods of traffic

5.5 Evolution under variable demand conditions

This chapter has clarified the impact of variable conditions to the evolution of the mean and variance of the queue. To complete this study a simulation of a peak hour is here presented and compared with the results computed with the Highway Capacity Manual delay function. The simulation is made under the assumption of deterministic capacity, keeping in mind that the conversion to stochastic capacity can be done by simply increasing the queue length of 20% the corresponding deterministic case.

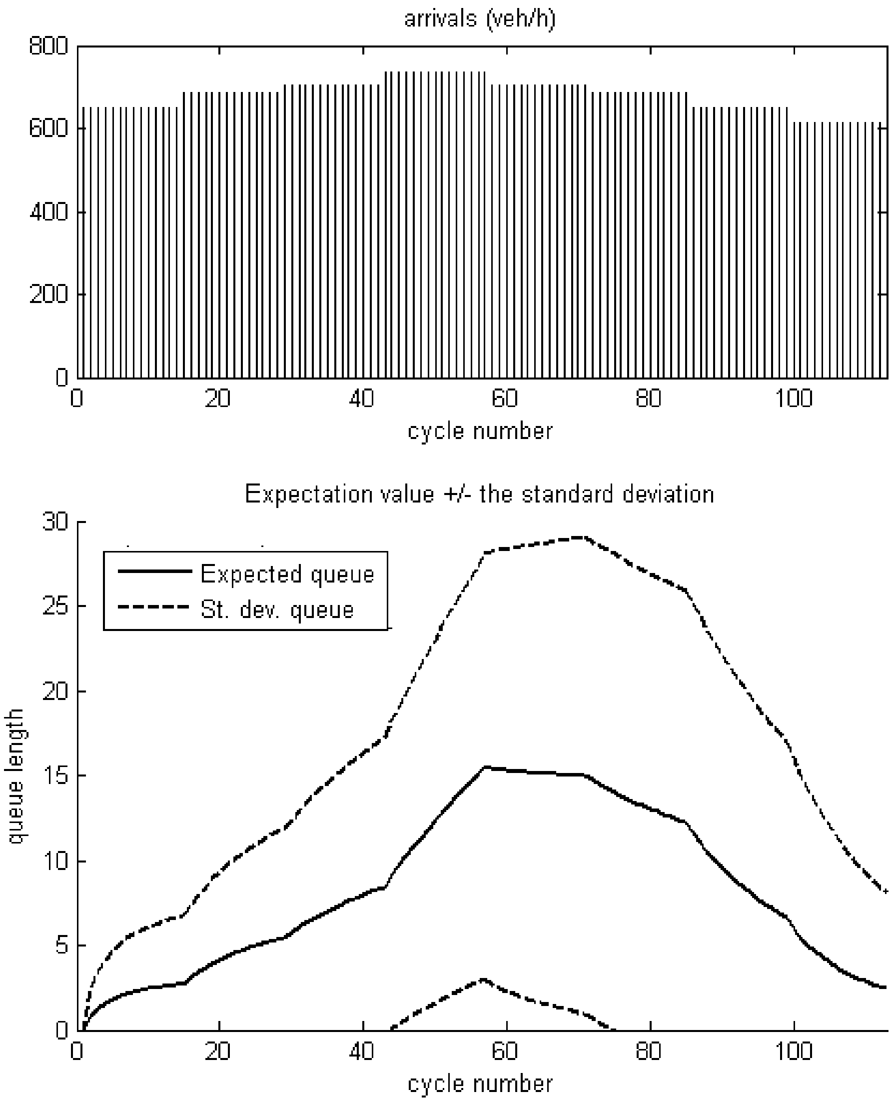

Figure 5.14 – Simulation of a peak hour, resultant queue length ± the standard deviation

The queue length does not show its maximum during the period of highest demand but at the end of the last oversaturated period (Figure 5.14). A static queuing model or any dynamic model, which does not properly include the effect of initial queues, cannot consider this behavior but it would predict the highest queue at the end of the highest volume-to-capacity ration. This would lead to and underestimation of the queue at the end of each subsequent period and consequently this error would be transmitted to the computation of vehicle delays.

5.6 Summary

Lack of a clear understanding of the relationship between the variability of the arrivals and the departures at a single signal motivated the analysis presented in this chapter. To operate this analysis the probabilistic approach presented in Chapter 4 has been applied. Such a method is valuable for two main reasons. It provides complete estimates of queues and delays by using aggregated data and due to this level of aggregation it enables one to compute these values with much higher computing speed than using microscopic simulation programs. On the other hand, this method can be used to generate a large dataset, which can be then used to derive empirical analytical formulas to solve long iterative processes like assignment problems or optimization problems.

This chapter has focused particularly on the analysis of the dynamic and stochastic aspects of the overflow queue in fixed-timed, isolated and single lane intersections. The case of an isolated signalized intersection has been analyzed under the whole range of possible demand conditions and in cases of stochastic capacity and deterministic or stochastic initial queue values. Analysis of the different behavior that can be observed under these assumptions has revealed that conditions of traffic in the neighborhood of saturation are strongly influenced by the random nature of demand and supply systems, creating an overflow queue, which can be enormously larger than uniform and incremental delay components described in the Highway Capacity Manual (HCM, 2000).

This method has quantified the error that affects standard analytical procedures like the HCM 2000 to compute vehicle delays especially when the model should compute decreasing queues after an oversaturated period, which frequently occurs during peak periods. All available methods up to date underestimate vehicle delays, or flatten delays within an evaluation period, instead of computing more accurately delays for each cycle. A new analytical method, which overcomes this problem, can be a valuable tool. This represents the main contribution of the next Chapter 6.

Combined analysis of average and standard deviation of the queue in time shows strong interdependence among these two characteristics, especially in saturated conditions of traffic. Such conditions of traffic create a delay that propagates in time and causes extra waiting times for vehicles approaching the intersection even several cycles after congestion had occurred. The ratio between standard deviation and variance influences the dynamic behavior of queues. This implies that an analytical expression for the standard deviation is also an important research issue.

6

Time-dependent models of overflow queues

6.1 Introduction

Available time-dependent functions provide expressions assuming an empty intersection at the start of the first red phase. Thus, only an increasing queue can be computed, both when intersections are oversaturated and when they are under-saturated. Akcelik's function (Akcelik 1980) is an example of such macroscopic queue models. As said in Chapter 3, no alternatives to Formula (4.1) have been proposed that model both increasing and decreasing overflow queues.

The opportunity to simulate the average and the standard deviation of the queue with the Markov model opens the opportunity to study various conditions of traffic. The behavior of the mean of the queue computed with the Markov model has been shown to be consistent with the results of the widely accepted models (i.e. (Akcelik 1980)) used in practice in the case of increasing queues with stationary demand conditions. Assuming the validity of the Markov model also in other conditions, namely non-zero initial queue, stochastic capacity, variable stationary demand conditions, a new macroscopic model for the dynamics of the queues is derived in this chapter. This new analytical model includes all these conditions in one simple and easy-to-use expression. The model presented in this chapter is able to represent faithfully the results of the average queue computed from the results of the Markov process in all conditions of traffic, both in assumption of uniform and non-uniform arrival rates. Moreover, having shown the important role of the standard deviation to the dynamic behavior of the expectation value of the queue, an analytical model for the standard deviation of the queue length is accordingly derived from the results of the Markov model.

The present chapter is structured as follows. Next section reviews the assumptions described in chapter 4 which apply also to the models derived in this chapter. Section 6.3 presents the model together with the procedure that led to its derivation and the calibration of its parameters. In Section 6.4 the same procedure is applied to derive an analogous expression for the standard deviation of the queue. Comparison between the proposed model, the Akcelik's model and the Highway Capacity Manual formula is presented in Section 6.5, showing the underestimation error one commits when a peak hour is simulated, especially in terms of decreasing queues. Comparison between the results in terms of delays gives analogous results.

6.2 Model assumptions

The linear deterministic function expressed by the Catling's formula (Catling 1977) describes satisfactorily the dynamic behavior of the expected value of queues in periods of severe congestion (volume to capacity ratio larger than 1.2) and it also well represents the behavior of decreasing queues when the queue has to recover from such a period and a large queue needs to be cleared, for example after the peak of a morning busy period. This formula is then a good starting point to derive the heuristic formula for time-dependent queues. For sake of clarity Catling's formula is here rewritten:

$$Q_t = \max\{Q_0 + (x-1)\cdot ct, Q_e\} \tag{6.1}$$

In under saturated and infinite steady-state conditions of the arrival distribution the behavior of the expected value of the overflow queue gradually changes from linear into an asymptotic behavior towards the steady-state equilibrium value, as shown in Chapter 5. A procedure to clearly visualize that the asymptotic behavior is sufficiently approximated by an exponential function as described in the next section. A simple method to combine these two trends is later proposed, together with a more simplified formula, which still satisfactorily follows the results of the Markov results.

It is worth recalling the assumptions, which bound the validity of the new analytical model as they did to the Markov model. The arrivals are restricted to be Poisson, thus different shapes of the arrival profile may result in different slopes for the behavior of overflow queues in time. Newell (Newell 1971) observed that the Poisson distribution is reasonable in cases of isolated intersections, but this distribution is no more valid when two or more intersections are closely spaced. Moreover the model is valid for single-lane sections and fixed controls. Implication in terms of average and standard deviation of the queue in such cases is discussed in chapter 8.

The important step in the introduction of this new model for the expected value of queues is the opportunity to model both the increasing and the decreasing behavior with one analytic formula, and the opportunity to estimate the dynamic propagation of the expected value of delays in time cycle by cycle. The contribution increases its importance if the same procedure to derive the heuristic model for the mean is also applicable to describe the standard deviation in the same broad conditions. Assuming the validity of the quadratic expression presented in chapter 5, formula (5.5), this expression represents the starting point to derive the heuristic formula for the standard deviation in time. Section 5.5 shows that the proposed model represents well the results of the Markov model also when a step-wise demand is assumed, guaranteeing continuity and smoothness in results in a dynamic process when the traffic conditions vary with time. A quantification of the error that affects the deterministic model in this context is reported, showing that this under estimation is made especially when the peak hour has for its large part a sub-period of slight under-saturation.

6.3 Time-dependent model for the expectation value

Stating the range of application of the linear behavior described by formula (6.1), the following analysis clarifies the behavior of such queues towards equilibrium when the demand distribution is assumed stationary for a very long evaluation period.

Recall also that in under-saturated conditions the queue is assumed to reach after some time an equilibrium value, which is well described by the formula suggested by Akcelik (Akcelik 1980):

$$Q_e = \begin{cases} \dfrac{1.5 \cdot (x - x_0)}{(1-x)} & \text{if } x \geq x_0 \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

where the degree of saturation x_0 represents the value above which equilibrium values are assumed a non-negligible. This value is represented by the following formula:

$$x_0 = 0.67 + \frac{c}{600} \tag{6.3}$$

where c is the assumed capacity (measured in vehicles per cycle).

As shown in Chapter 5 the linear behavior satisfactorily represents the expected value of the queue length only for some cycles and this interval of time depends on the value of the initial value and the coefficient of variance. The behavior towards equilibrium assumes gradually an asymptotic curve, and the equilibrium value is met only after several cycles, depending on the degree of saturation. An expression for this transition phase is derived in the following section. This expression should be also able to guarantee continuity, in order to guarantee its applicability in problems involving iterative processes (e.g. optimization problems, assignment processes). The derivation of this expression starts from considering the case of decreasing queues and analyzing the behavior towards the equilibrium value. An extension of this formulation to increasing queues follows.

6.3.1 Derivation of the exponential evolution

Figure 6.1 shows how the expectation of the overflow queue decreases from an initial deterministic value of 40 vehicles together with the growth of the standard deviation, for a volume-to-capacity ratio of $x = 0.9$. If the initial queue is large and the queue is decreasing, the standard deviation grows to a high value and when the queue becomes close to the standard deviation, the latter decreases and goes also to an equilibrium value.

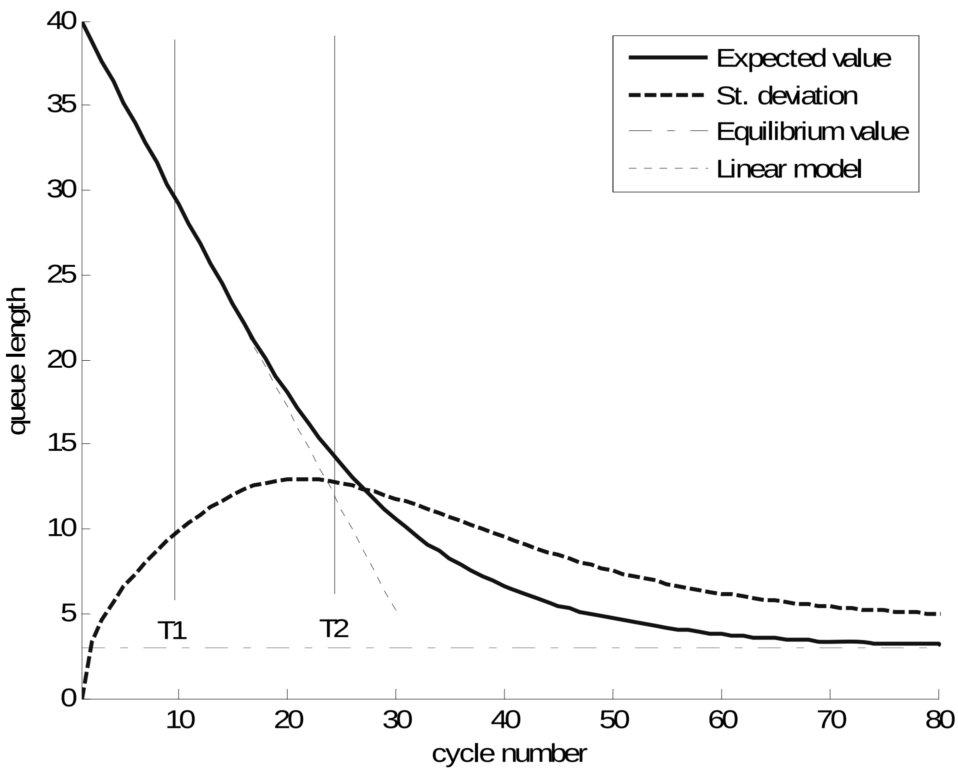

Figure 6.1 – Evolution of mean and standard deviation of queues starting from Qo=40 with x=0.90

If one computes the logarithm of the derivative of the expectation value, a linear trend for the behavior after a certain number of cycles is observed, as it is shown in Figure 6.2.

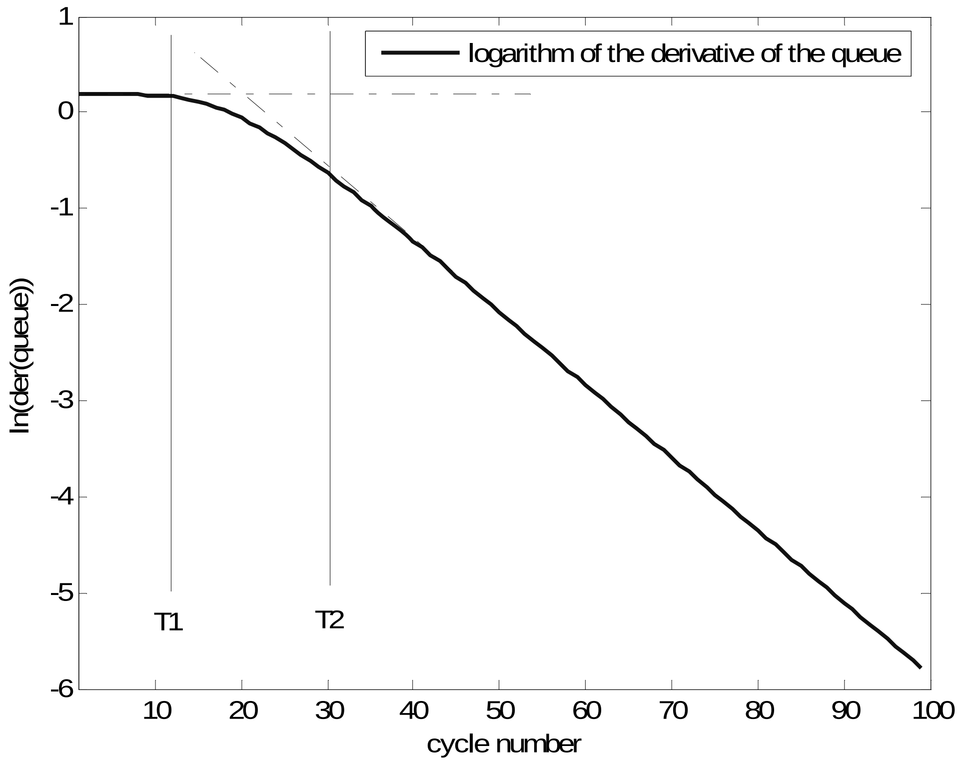

Figure 6.2 – Logarithm of the derivative of the expected value for *x*=0.90 and Qo=40

The logarithm of the derivative of the queue in time (expressed in cycles) can be therefore approximated by a linear function:

$$\ln\left(\frac{\partial Q(\mathbf{y},t)}{\partial t}\right) = \beta(\mathbf{y})t + \gamma(\mathbf{y}) \tag{6.4}$$

The vector $\mathbf{y}$ represents the vector of state variables (number of arrivals, signal capacity, signal settings) that influence the value of the parameters β and γ. These variables will be specified later in this chapter. An expression of the asymptotic behavior of the queue towards equilibrium can be derived by solving the first-order differential equation (6.4). The approximate expression of the expected value of the queue length is thus found by computing the following integral:

$$Q(\mathbf{y},t) = \int_{T2}^{\infty} e^{\beta(\mathbf{y})\tau + \gamma(\mathbf{y})} d\tau \tag{6.5}$$

Assuming that $\lim_{t\to\infty} Q(\mathbf{y},t) = Q_e$ with Q_e computed with formula (6.2) being the value at equilibrium, the queue behavior towards equilibrium follows the expression described by Formula (6.6):

$$Q(\mathbf{y},t) = Q_e + \gamma'(\mathbf{y}) \cdot e^{-(\beta(\mathbf{y})\cdot t)} \tag{6.6}$$

where

$$\gamma'(\mathbf{y}) = \frac{e^{-\gamma(\mathbf{y})}}{\beta(\mathbf{y})} \tag{6.7}$$

This asymptotic behavior describes the evolution of the expected value of the overflow queue length after a certain number of cycles for both increasing and decreasing cases.

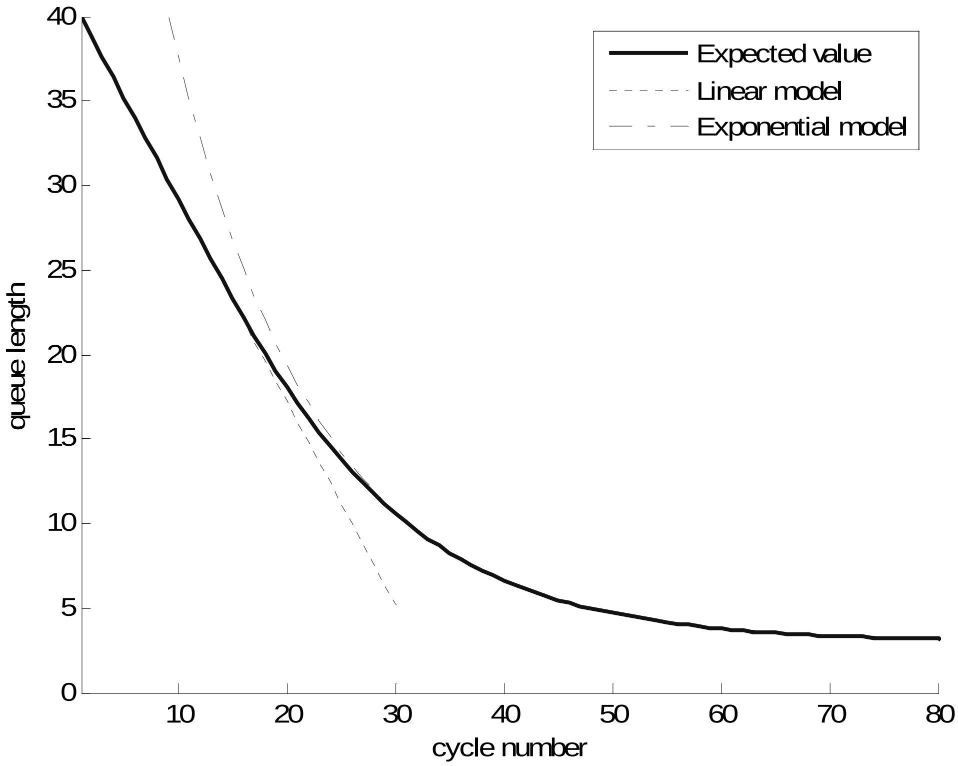

Figure 6.3 –Queue, linear initial trend and exponential asymptotic behavior for x=0.90 and Qo=40

The behavior of the logarithm of decreasing queues, as shown in figure 6.1, is initially linear until time T1, then follows a transient phase and gradually modifies its behavior towards the exponential trend at T2. Figure 6.3 displays the queue computed in the example in Figure 6.1 together with the linear and the exponential functions that approximate the first and the last areas subdivided by T1 and T2.

In the case of increasing queues the logarithm of derivative of the queue length shows, after some cycles, a linear behavior, whose slope is equal to the analogous decreasing case, while the constant γ' will change. The case of increasing queues does not show the linear initial behavior since no deterministic queue can be observed but only a transient phase will precede the exponential behavior in such conditions.

6.3.2 The three-phase model for the decrease of overflow queues

The transient state is approximated by introducing a time-dependent weight $\alpha(\mathbf{y},t)$ to the linear Q_{lin} and the exponential trend $Q_{\exp}$:

$$\alpha(\mathbf{y},t)\cdot Q_{lin}+[1-\alpha(\mathbf{y},t)]\cdot Q_{\exp} \tag{6.8}$$

The transient weight function varies in the domain $[T1,T2]$ and maps in the co-domain $[0,1]$, it is decreasing and satisfies the following conditions:

$$\left.\frac{\partial\alpha(\mathbf{y},t)}{\partial t}\right|_{T1}=\left.\frac{\partial\alpha(\mathbf{y},t)}{\partial t}\right|_{T2}\simeq 0 \tag{6.9}$$

The above conditions guarantee that the transient phase extends the linear and the exponential functions with continuity together with their first derivatives. Section 5.3.3 describes the method to derive an approximate expression for the weight function. Summing up, the behavior of decreasing queues starting from a deterministic initial value is expressed by the following three-phase model:

$$Q(\mathbf{y},t)=\alpha(\mathbf{y},t)\cdot[Q_0+(x-1)\cdot c\cdot t]+(1-\alpha(\mathbf{y},t))\cdot[Q_e+\gamma'(\mathbf{y},t)\cdot e^{-\beta(\mathbf{y},t)\cdot t}] \tag{6.10}$$

In the following of the thesis this model will be referred to as the *Van Zuylen-Viti formula*.

6.3.3 Calibration of parameters

In order to compute analytically the overflow queue expressed by Formula (6.10) one needs to have a closed form expression for the transient function α and for the parameters β and γ'. The parameter β controls the curvature of the exponential function. Figure 6.4 shows the behavior of the logarithm of the derivative of the queue with respect to different degrees of saturation and different initial queues.

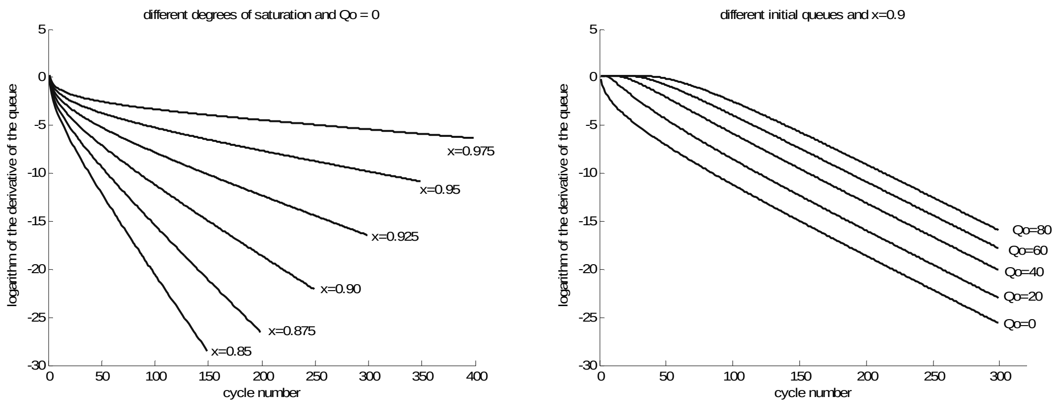

Figure 6.4: Behavior of the logarithm of the overflow queue with different degrees of saturation and different initial queues

Analysis of overflow queues for different degrees of saturation, cycle lengths and initial queues show that the parameter β depends only on the first state variable and it can be expressed by the following formula:

$$\beta = \frac{(1-x)^2}{0.2} \tag{6.11}$$

The parameter γ controls the position at which the exponential follows after the linear and the transient phases. The higher the initial queue value, the later in time the exponential state is expected to occur. Analysis of the Markov data shows that this parameter depends also on the degree of saturation and it can be approximated by the following formula:

$$\gamma = (1-x) \cdot Q_0 + 1.5 \tag{6.12}$$

The function $\alpha(\mathbf{y},t)$ should be a function that assumes the value 1 until the moment that the standard deviation is about 50% of the expected queue value. The function should become zero when the extrapolated linear decrease arrives at the zero queue length. An example of $\alpha(\mathbf{y},t)$ is given in Figure 6.4, obtained by solving the simple problem of finding the value of a of the equation below, for any value of t:

$$Q_{MC} = a \cdot Q_{lin} + (1-a) \cdot Q_{\exp} \tag{6.13}$$

where Q_{MC} is intended the result of the Markov simulation.

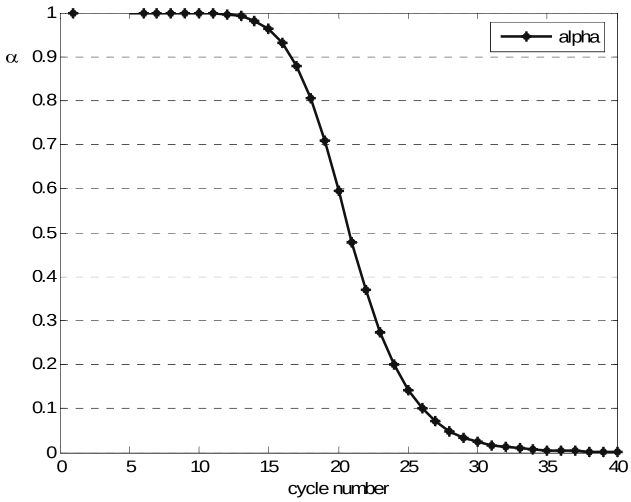

Figure 6.5 – Weight function alpha for x=0.90 and Qo=40

The transition phase in the example starts after around 10 cycles, while the exponential behavior is dominant after 30 cycles. A reasonable approximation of the function $\alpha(\mathbf{y},t)$ is for example a logistic function:

$$\alpha(\mathbf{y},t) = \frac{1}{1+e^{\mu(\mathbf{y},t)(t-T_0(\mathbf{y},t))}} \tag{6.14}$$

where $T_0(\mathbf{y},t)$ is the time that the standard deviation is equal to the expectation value of the queue length and $1/\mu(\mathbf{y},t)$ is approximately equal to the time between the time that the standard deviation becomes equal to 50% of the expected value of the queue, T1, and the time that the standard deviation is approximately equal to the expected value of the queue length, T2, (as seen in Figure 6.1).

Analysis of the influence of the state variables shows that the duration of the transition phase depends only on the degree of saturation. The parameter μ can be approximated by the following formula:

$$\mu = \frac{0.1 \cdot x}{(1-x)^2} \tag{6.15}$$

The parameter T_0 controls the position in time of the transition phase, and it is both dependent on the degree of saturation and the initial queue length. An approximation of this parameter is given by the following formula:

$$T_0 = \frac{0.15 \cdot Q_0 \cdot x}{2(1-x)} \tag{6.16}$$

6.3.4 Simplified bi-phase model

Formula (6.10) can replace the Catling's Formula (6.1) and it has also the property of describing the non-linear behavior described by several models presented in chapter 3, among others the Akcelik's Formula (3.31). However, its expression can be simplified with, as it will be shown, little extra approximation. Formula (6.6) can in fact be rewritten equivalently in the following form:

$$Q_{\exp}(t) = Q_e + (\bar{Q} - Q_e) \cdot e^{-\beta t} \tag{6.17}$$

obtained by simply substituting γ' with an equivalent parameter, $(\bar{Q} - Q_e)$.

The simplified formulation can be derived by considering simply the exponential function extended by the linear function in the point in which the latter equals the derivative of the exponential. Analyzing the parameters β and γ', the first determines the curvature of the exponential function while the second determines the position. If one considers the parameter β determined with formula (6.11) while γ unknown, he can determine the point Q^* where the derivative of the exponential equals the deterministic function by a simple mathematical problem: find the value of Q^* at time t^* that respects the system of equations:

$$\begin{cases} Q_{linear}(t^*) = Q_{\exp}(t^*) \\ \left.\dfrac{\partial Q_{linear}}{\partial t}\right|_{t^*} = \left.\dfrac{\partial Q_{\exp}}{\partial t}\right|_{t^*} \end{cases} \tag{6.18}$$

where

$$\begin{aligned} Q_{linear}(t) &= Q_0 + (x-1)ct \\ Q_{\exp}(t) &= Q_e + (Q^* - Q_e) \cdot e^{-\beta t} \end{aligned} \tag{6.19}$$

The value of Q^* is thus found by solving the system of equations under the variable t. It is possible to show that the system of equation leads to the equation:

$$\beta t - \frac{1}{e^{\beta t}} = \kappa \tag{6.20}$$

where

$$\kappa = \frac{\beta(Q_0 - Q_e)}{(x-1)c} \tag{6.21}$$

Equation (6.20) has a closed form solution:

$$t^* = \frac{lambertW(e^{\kappa}) - \kappa}{\beta} \tag{6.22}$$

where $lambertW$ is the Lambert or Omega function. If t^* is positive the solution is applicable and the value of Q^* can be then found substituting this value in the linear function, otherwise the linear part is neglected and the queue assumes the expression:

$$Q(t) = Q_e + (Q_0 - Q_e) \cdot e^{-\beta t} \tag{6.23}$$

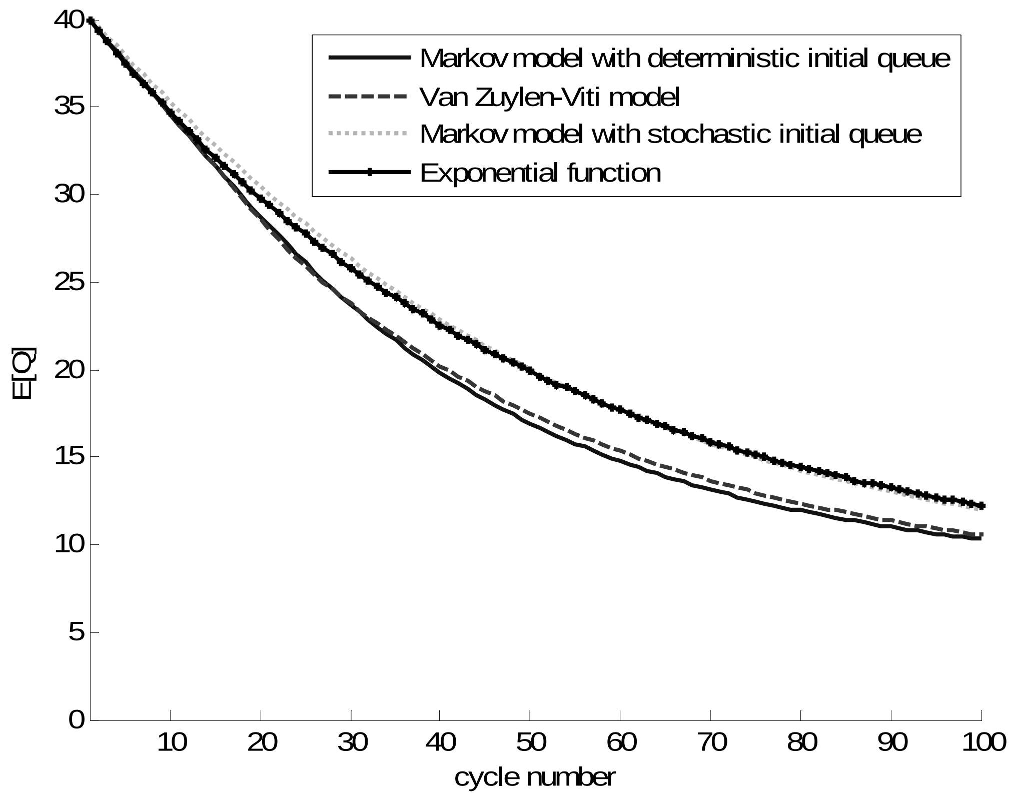

Figure 6.6 – Comparison between Markov results and heuristic formula for decreasing queues

Figure 6.6 shows an example of overflow queue computed with $x = 0.95$ and initial value of 40 vehicles in comparison with the result of the simplified formula. Visual inspection clearly shows the good fit of the approximate expression.

6.3.5 *Extension to increasing queues*

Formula (6.23) can also well approximate increasing queues, both starting from a zero initial queue and with an initial queue that is smaller than the equilibrium value. Figure 6.7 compares the Markov simulation results with the results of applying Formula (6.23).

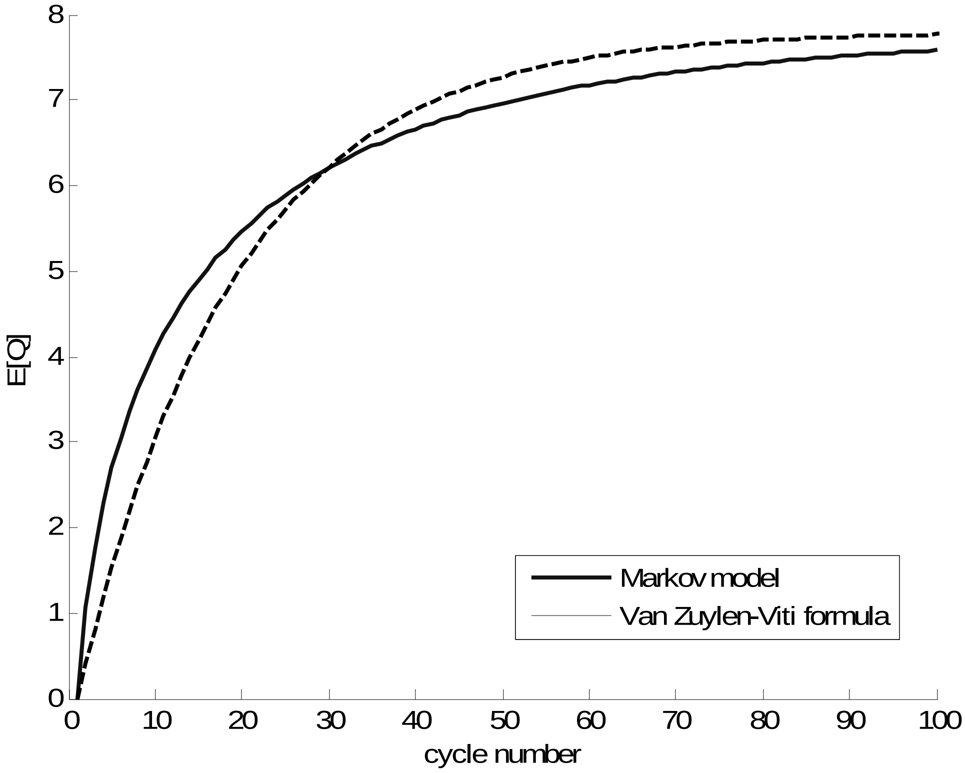

Figure 6.7 – Comparison between Markov results and heuristic formula for increasing queues

Apparently, the approximate formula works better when the queue is not zero. This might be explained by the stronger influence of the transition phase in this condition. One may consider to slightly modify the value of the parameter β in order to obtain a better representation of this case.

In conclusion, the dynamics of the overflow queue is represented both in the decreasing and in the increasing case by analogous expressions. The average overflow queue represented is continuous and differentiable within each period of stationary demand conditions and fixed control settings, while it is still continuous if computed with non-uniform arrivals. This finding confirms the importance of such heuristic method in problems, which require continuous Dynamic Loading Processes or involves optimization methods.

6.3.6 Behavior under variable demand conditions

To test the goodness-of-fit of the Van Zuylen-Viti formula is compared with the Markov model and the Catling's (Catling 1977) formula in a simulation of a peak period. The example of figure 5.14 computed with the mesoscopic model is shown again in Figure 6.8 together with the two analytic expressions.

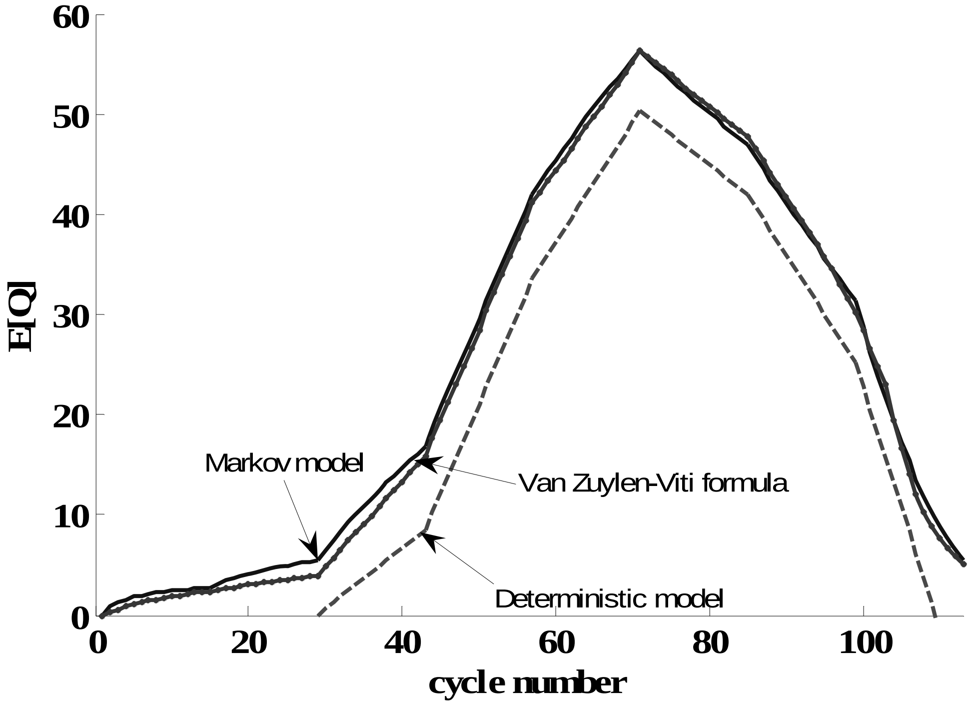

Figure 6.8 – Comparison between Markov results, Catling's and heuristic formula in a peak hour

The heuristic formula follows faithfully the results of the Markov chain, while the Catling's expression underestimates the queuing evolution during the whole evaluation period, since it does not compute an average overflow queue for the first periods and it increases the error at the tail of the evaluation period. This underestimation of overflow queues causes an underestimation of expected delays. Section 6.5 shows how much this error affects the result of analytical models most frequently used in practice.

6.4 Time-dependent model for the standard deviation

Under the assumption of deterministic initial values, the standard deviation of the queue starts evolving according to formula (5.5). One more time, the queue displayed in Figure 6.1 is used as example to derive an expression to the standard deviation that is analogous to the one for the expectation value expressed in formula (5.5). Computation of the logarithm of the derivative of the standard deviation shows also a linear behavior after a certain number of cycles and the slope is the same as in the case of average overflow queues. Thus, this part is well approximated by formula (6.4). Figure 6.9 compares the standard deviation computed for $x = 0.90$, initial queue of 40 vehicles and for 100 cycles with the correspondent exponential and quadratic first and third phases. It looks also that the three phases of the standard deviation are delimited according to ones defined for the mean.

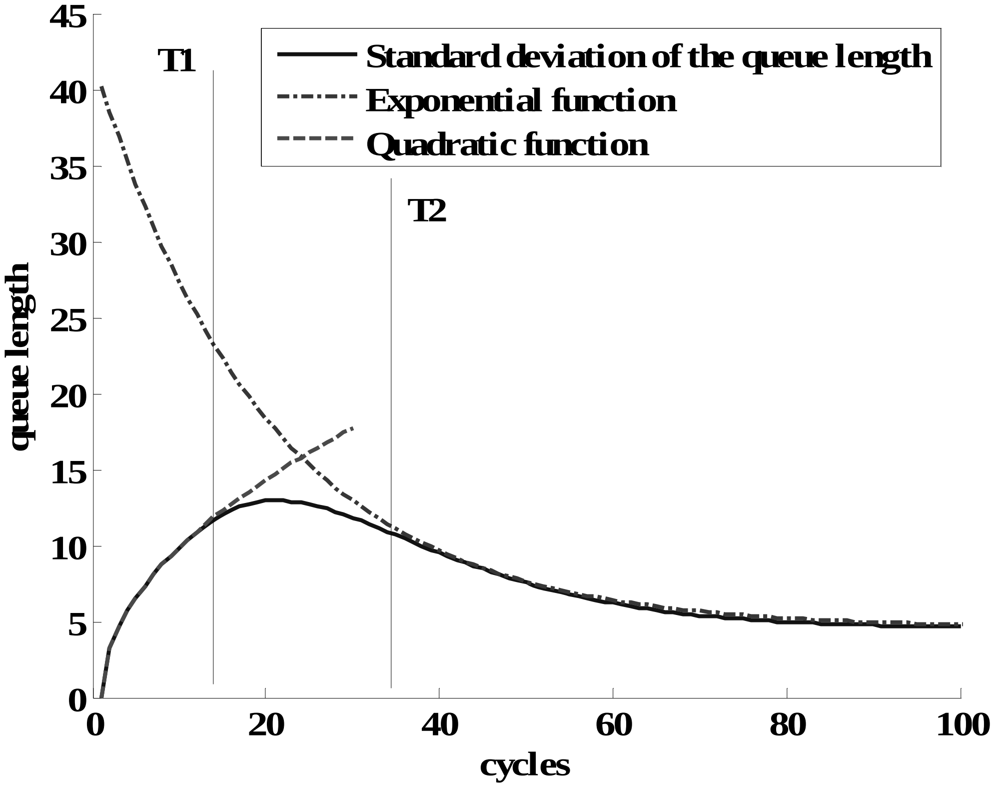

Figure 6.9 – Standard deviation, quadratic and exponential approximations

The quadratic behavior, in the assumption of deterministic non-zero initial values, increases, but after few cycles it starts decreasing and later on a transition phase precedes an exponential evolution. This suggests the introduction of a three phase model for the standard deviation in the same manner the evolution of the average was derived.

Formula (5.5) was derived in the case of zero initial value. The quadratic form follows a slightly different behavior, and the position in time of the maximum value depends on the length assumed for the initial value. It is assumed that the initial condition follows the form:

$$\sigma\{Q\} = at^2 + bt + c \tag{6.24}$$

Since the initial value is assumed to be zero then $c = 0$. Different initial values show that the initial evolution for the first cycles does not change. This finding suggests that the parameter b depends only on the degree of saturation and not on the initial queue length. Analysis of different degrees of saturation shows that a good approximation of b can be obtained with the following expression:

$$b = \frac{1}{2}\sqrt{\frac{xc}{t}} \tag{6.25}$$

The parameter a can be derived by considering the value of the maximum for different initial values and different degrees of saturation. The maximum, according to the Expression (6.25) is given at the value $t = b^2 / 2a$. The following expression approximates the parameter a:

$$a = \frac{1}{2} \cdot x \cdot c \cdot t \cdot \sqrt{\frac{2.4 \cdot (1-x)}{Q_0}} \tag{6.26}$$

The following formula extends formula (5.5) to include cases of non-zero queues:

$$\sigma[Q] = \frac{1}{2} \cdot x \cdot c \cdot t^3 \cdot \sqrt{\frac{2.4 \cdot (1-x)}{Q_0}} + \frac{1}{2} \cdot \sqrt{x \cdot c \cdot t} \qquad \text{for } t < T1 \tag{6.27}$$

If the initial value is assumed stochastic the initial quadratic behavior does not occur and the behavior of the standard deviation can be sufficiently described by simply the exponential expression:

$$\sigma\{Q\} = \sigma\{Q_e\} + (\sigma\{Q_0\} - \sigma\{Q_e\}) \cdot e^{-\beta \cdot t} \tag{6.28}$$

where $\sigma[Q_e]$ is the value at equilibrium computed by using formulas (5.3)-(5.4).

Figure 6.10 shows the behavior computed with the Markov chain process in comparison with the exponential function. The slight difference during the first cycles is caused by the assumed Normal distribution for the starting value with standard deviation equal to the average value.

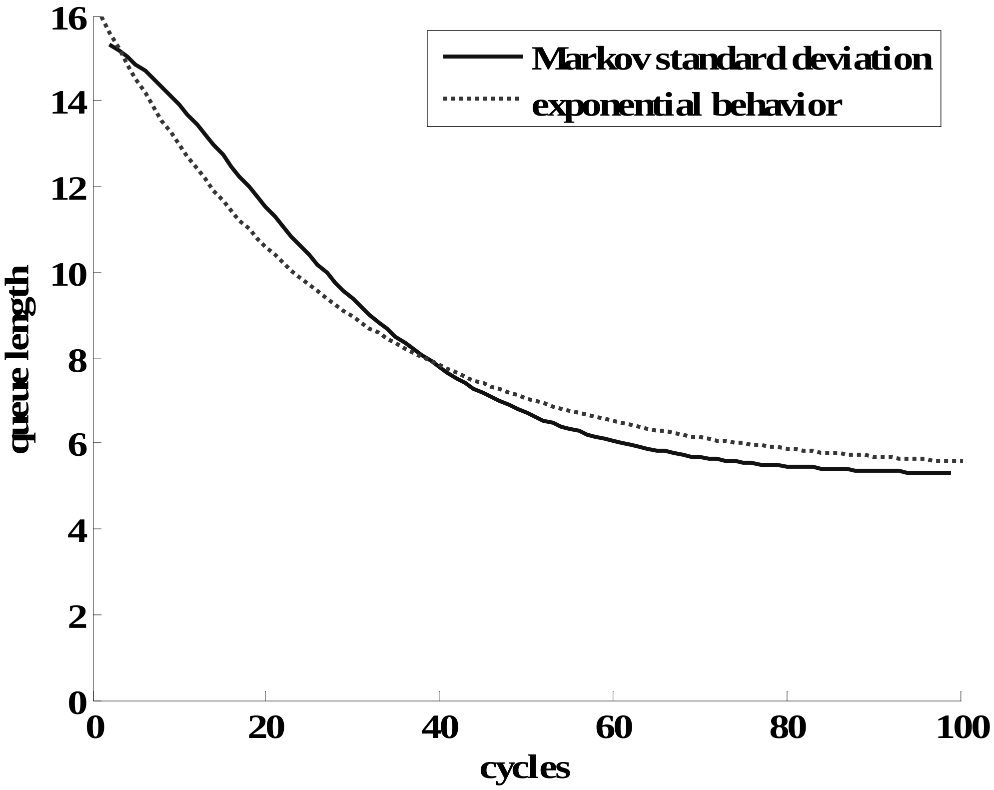

Figure 6.10 – Comparison between Markov results and exponential function for Qo=16 and x=0.90

These considerations suggest a general expression for the standard deviation, which is similar to the one introduced for the average value.

6.4.1 Three phase model

The analogy with the three phase behavior of the average overflow queue suggests the introduction of a similar formulation for the standard deviation. If the transient phase is represented by the same form of Formula (6.10) the model for the standard deviation can be shown to be approximated by using the same weight function obtained by using the alternative formulation of the exponential function (6.17):

$$\sigma\{Q(\mathbf{y},t)\} = \alpha_{SD}(\mathbf{y},t)\cdot Q_{quad} + (1-\alpha_{SD}(\mathbf{y},t))\cdot Q_{\exp} \tag{6.29}$$

The function $\alpha_{SD}(\mathbf{y},t)$ is slightly different from α. The functional form can be still satisfactorily approximated by the logistic function in Formula (6.14). The duration of the transition phase does not change significantly, thus Formula (6.15) for the parameter μ is also applicable in this context. The expression for the duration of the transition phase T_0 can instead be approximated by multiplying Formula (6.16) with the multiplicative factor approximated with Formula (5.4).

6.4.2 Simplified bi-phase model

The expression of the standard deviation introduced in Formula (6.29) can be simplified in the same way it was done for the expression of the mean. The transition phase can thus be neglected with a reasonably small payoff in terms of model accuracy if the following system of equation is solved:

$$\begin{aligned} & Q_{quad}(t^*) = Q_{\exp}(t^*) \\ & \left.\frac{\partial Q_{quad}}{\partial t}\right|_{t^*} = \left.\frac{\partial Q_{\exp}}{\partial t}\right|_{t^*} \end{aligned} \tag{6.30}$$

The system of equations provides the value of t and the value of the standard deviation in which the quadratic and the exponential functions have the same value and the same first derivative. The solution does not have a closed form expression as in Formula (6.22) but it is solvable numerically. Figure 6.11 compares the results of the Markov model and the simplified analytical formula for the standard deviation computed with Formula (6.30).

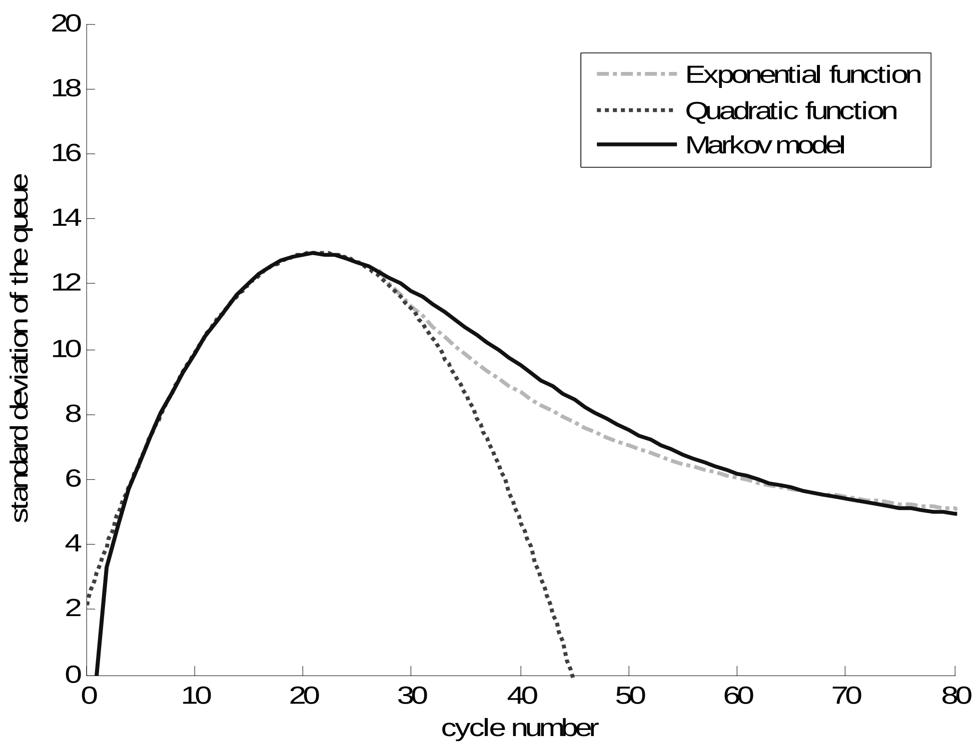

Figure 6.11 – Comparison between Markov results, and simplified model for Qo=40 and x=0.9

6.4.3 Behavior under variable demand conditions

For sake of completion the standard deviation of the overflow queue in the example of Section 5.3.4 is here presented and compared with the presented empirical model.

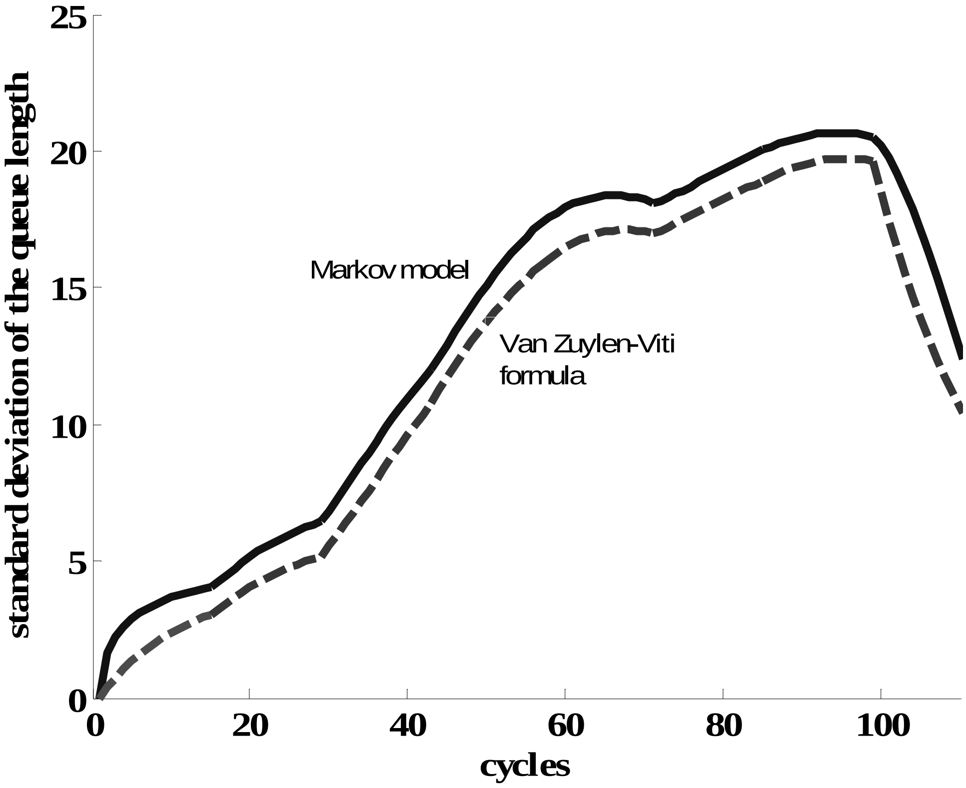

Figure 6.12 – Comparison between Markov results, and simplified model for a peak hour

It looks that the empirical model underestimates the model computed with the Markov chain. This error is produced only during the first period, where the results of the mesoscopic model are computed under the assumption of zero average and standard deviation of the overflow queue. The underestimation error is although always smaller than 2 vehicles even for oversaturated conditions.

6.5 Comparison with the HCM 2000 delay formula

The estimation power of the models proposed in this chapter can be quantified in terms of average delay. The Highway Capacity Manual (TRB 2000) control delay formula represents one of the most frequently used set of formulas practitioners use in planning, design and evaluation of travel times in transportation networks. There is no specific computation of the variability of the delay in the manual, but only an average delay formula is provided, as presented in Section 3.3.4. The control delay expressed in this and all other official manuals evaluates the average vehicle delay for a fixed analysis period, thus this value is uniformly distributed during the whole period. Moreover the manuals assume stationary demand conditions within this period.

The computation of cycle delay as described in Section 4.6 allows the traffic analyst to compute the vehicle delay according to the evolution of the queue cycle by cycle, as computed with the Markov model and with the empirical formula. The Markov and the Van Zuylen-Viti models give also the opportunity to represent the delay as continuous functions, while the one computed with the HCM2000 is step-wise. From the distribution of the queue length one can also derive the probability distribution of the delay a traveler may occur. In Chapter 2 it has been observed that this information is very important if one includes travelers' response to congested and variable conditions. The benefits of these approaches for example in a Dynamic Traffic Assignment are thus evident.

The HCM2000 subdivides the control delay in three components, as seen already in Chapter 3. The first component represents the effect of a vehicle to arrive during the effective green or red periods. The queue in front of the signal in this condition is always assumed zero. The random component is instead related to an overflow queue to occur, and it computes the extra delay the temporary overflow occurred in a cycle can produce. Finally the third component evaluates the extra delay an initial queue value produces.

Table 6.1 – Assigned demand and resultant average delay with Markov chain, model and HCM

Average delay	**Flow (veh/h) for 15 minutes intervals**							
	648	**684**	**756**	**828**	**792**	**684**	**648**	**540**
HCM	22.42	23.66	53.74	113.02	105.89	79.07	63.05	37.77
Q_{MC}	25.32	36.47	70.99	156.91	243.28	246.84	189.36	84.83
Q_{model}	28.19	40.86	74.18	161.89	260.60	278.07	201.24	98.05
$\sigma\{Q_{MC}\}$	28.72	43.94	83.06	171.43	255.82	260.98	210.74	109.52
$\sigma\{Q_{model}\}$	34.71	46.73	62.90	85.66	102.26	121.58	130.24	114.62

The models proposed in Chapter 4 and the Van Zuylen-Viti formulas compute the overflow queue, which might occur in a period of either stationary, or non-uniform arrivals. The computation of delays using these models and the formulas presented in section 4.4 is equivalent to computing both second and third components. For this reason the results below are only referred to these components. Table 6.1 compares the results of delays derived from the queue evolution presented in the example in Figure 6.8 for the expected value and figure 6.12 for the standard deviation measures.

For the first two periods the delay computed with the Highway Capacity Manual is only consisting on the incremental delay, while no initial queue delay is considered. Initial queue delay is computed with the Catling's Formula (6.1). Looking at Figure 6.8 the deterministic function starts increasing only during the third period and it clears in the middle of the last period. The overflow computed with the Markov and the Van Zuylen-Viti models have already a positive average value during the first two periods and it does not clear completely the queue at the end of the last period. This positive residual queue produces delay also outside the total evaluation period. The method described in section 4.4 computes the extra delay one has to consider for all vehicles entering in one period, since a part of them may be served only after several cycles or even at the end of the evaluation period.

Table 6.1 shows the enormous difference between the average delays computed with the HCM2000 and the Markov model of queues combined with the computation of cycle delay. The difference is particularly evident in the period of decreasing queues, when the HCM2000 computes a smaller delay for the entering vehicles, while a consistent part of the flow, which has entered the system during the growing-up part of the queuing process, still has to be served and creates a delay, which is larger than the one associated to the peak flow. In the decreasing part of the peak period the delay computed with the Markov model is over 3 times the one computed with the HCM2000. A small fraction of this difference is due to the consistent underestimation of the queue length computed with the Catling's formula, while a large part is due to the different method for computing the expectation value of the delay, especially when there is an initial queue delay component.

In all periods the standard deviation assumes a value approximately equal to the average. These findings give a qualitative impression about the uncertainty in the value of delays in such peak periods, and how uncertain is the time these delays propagate involving off-peak periods. The standard deviation appears to be underestimated by the heuristic model. Since there was very small difference in terms of standard deviation of the queue, this error is due to the use of average queues and arrivals in the computation of cycle delays as described in Section 4.6.

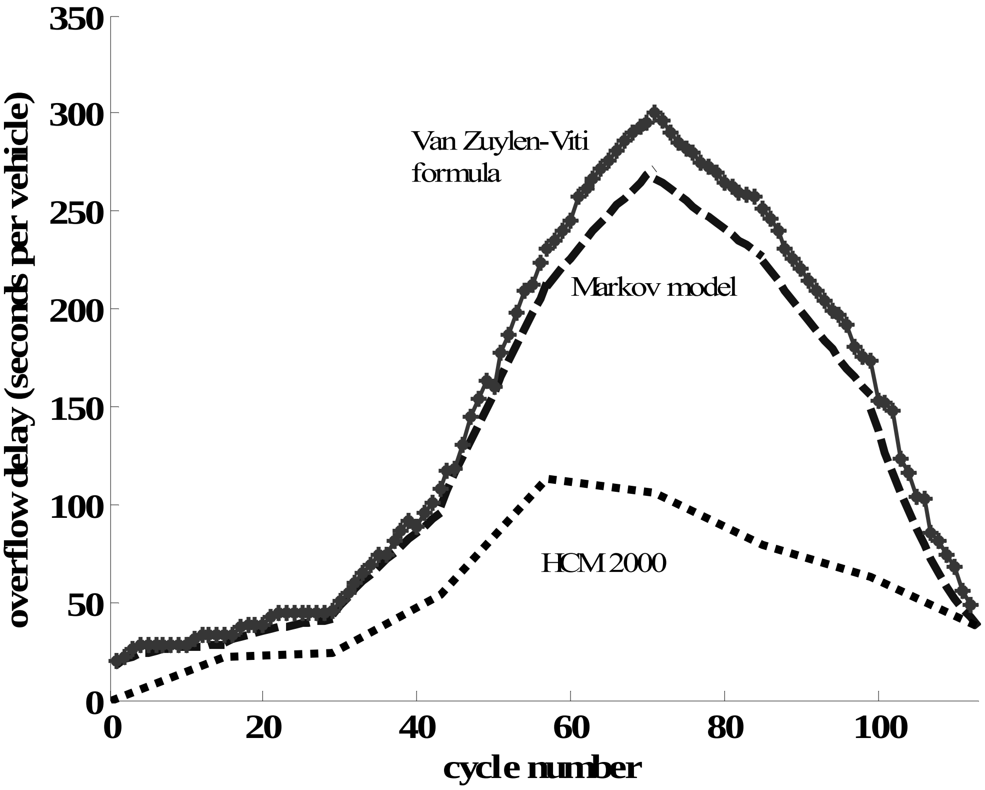

Figure 6.13 – Comparison between delays from Markov, heuristic and HCM models for a peak hour

Comparing the vehicle delay computed with the analytical queue model and the method described in Section 4.4 with Markov and HCM models it seems that the results of the Van Zuylen-Viti models are much more consistent with the Markov model results. Figure 6.13 compares the results of the empirical formula with the ones computed with Markov and HCM models.

6.6 Summary

Traffic practitioners agree that no valid queuing formulas exist which are general enough to be used in evaluation or prediction problems and in several transportation problems like assignment processes and optimization problems. Available macroscopic models of overflow queues and delays have assumptions which limit their estimation power in dynamic scenarios.

This chapter has presented a new set of analytical formulas for the expectation values of queues and delays at isolated signalized intersections. The data simulated with the Markov chain process have been used for this. A linear decrease followed by an exponential behavior has been found when queues recover from large initial values and the signal operates near capacity. This finding has suggested a new formulation for the dynamics of the overflow queue, which combines the deterministic linear behavior with a smoother asymptotic behavior towards the equilibrium value.

Heuristic functions have been derived for the expectation value and the standard deviation of the queue in time. These models have a broader area of use than official manuals as for example the Highway Capacity Manual 2000, since they reproduce the expected evolution of queues and their variability as function of time, without the necessity to fix an evaluation period but they provide estimates for every cycle. Their easy formulation makes them appealing in design and planning problems.

The models proposed can compute queues and delays assuming both uniform and non-uniform arrivals. This feature makes them suitable for Dynamic Loading processes, but also to make short-term prediction of expected waiting time. Given the large difference between the results of the manuals and the ones computed with the Markov model described in chapter 4 and the models proposed in this chapter, delays computed with the manuals strongly underestimate the delays produced by congested conditions, especially fail in evaluating the dynamic propagation of delays forward in time and the consequences a congested period causes to off peak periods.

7

Consistency between probabilistic models and microscopic simulation

7.1. Introduction

The analytical Van Zuylen-Viti queue models presented in Chapter 6 have the property to describe the dynamic evolution of the overflow queue and its temporal distribution with two simple analytical expressions. These models enable the traffic analyst to evaluate the stochastic effects of demand fluctuations especially in conditions near capacity and to model the transition between low degrees of saturation to large ones and vice versa with a continuous function. The model has been derived from taking the expected value and the standard deviation of the results of the Markov model, which also approximates the evolution of queues by computing the probability of each overflow queue state from the distribution of arrivals and departures at each time step and from the queue length distribution at the previous time step.

One may question whether this method is a valid representation of reality or not. The traffic system is a complex combination of physical and behavioral mechanisms, while the Markov model is simply a combination of mathematical relationships between probability distributions. Validation of the analytical and the Markov models should be made using field data. It is rather difficult acquire such dataset, since several days of observations should be made in order to make valid estimates of mean and standard deviation of queues. Even if one can collect the queue length dynamics for several days it is rather unlikely that one can observe periods of stationary demand conditions, which are long enough to observe equilibrium conditions for the average overflow queue length. Moreover, collection of such an amount of real data constitutes an enormously expensive job in terms of time and costs. Teply (Teply 1989) stated that, even after collecting such a large amount of field data, a perfect match with any model can never be expected.

A microscopic simulation program may represent in this case alternative to field data. Therefore, this chapter compares the results of the Van Zuylen-Viti analytical formulas and the Markov model with simulations computed with a microscopic program, VISSIM (PTV 2003). This simulation has been done for two situations: the isolated intersection and the intersection in a network.

To do so, Section 7.2 introduces the microscopic simulation technique and explains the property of microscopic models to catch the dynamic and stochastic aspects of flows propagating on a network. Section 7.3 presents the scenario analyzed using the microscopic software program VISSIM. Section 7.4 analyses the distribution of the simulated overflow queues under stationary demand conditions. Section 7.5 compares the simulated data with the results of Markov and proposed models in various conditions of traffic in an isolated, single-lane intersection. Section 7.6 analyzes the influence of an upstream signal and finally Section 7.7 gives a synthesis and points out the main conclusions of this chapter.

7.2. Queue comparison with microscopic simulation

Microscopic simulation programs are very powerful tools used in several areas of the transportation theory and practice. The Federal Highway Administration considers microscopic simulation to be a very effective technique in the traffic analysis, since “this approach allows one to evaluate heavily congested conditions, complex geometric configurations, and system-level impacts of proposed transportation improvements that are beyond the limitations of other tool types. However, these models are time consuming, costly, and can be difficult to calibrate” (FHWA 2006). Pursula (Pursula 1999) recognises microscopic models to be applicable in all fields of the traffic engineering.

These programs can be used, and have been used in past studies, to validate analytic models and assess their accuracy since Webster (Webster 1958). In his seminal paper Webster already pointing out the complexity of validating delay models through direct observation since variations in traffic are uncontrollable. He therefore decided to use “a method whereby the events on the road are reproduced in the laboratory by means of some machine which simulates behavior of traffic”. Akcelik (Akcelik 2001) discusses the pros and cons of this approach in the validation of analytical formulas, giving appreciation to this method for the way it models the complexity and uncertainty characterizing the urban traffic. On the other hand he warns on a misuse of this approach since it is still a representation of the reality, therefore conclusions among the mutual accuracy and reasonableness of models can be done, but no conclusion can be done whether a model is better than another.

Since Webster few studies have considered the large variation that can be calculated in the simulation results. Tian (Tian 2002) analyzed the problem under various conditions of traffic, stressing the importance of using a sufficient number of simulation runs to estimate the variations of the delay. Both Tian and Dion (Dion 2005) agree in pointing at the largest variation and disagreement between the available analytic models and microsimulation when the degree of saturation is near one. Moreover Tian finds disagreement also between different microscopic programs in evaluating the variability of delays, especially when a link is highly saturated. Despite these issues, microscopic programs still remain necessary to evaluate lower level-of-detail models.

Several random simulations are needed to give accurate estimations of the expected value of overflow queues and their standard deviation. Moreover, several different traffic conditions had to be analyzed in order to assess the consistency of the microscopic results with the model proposed in chapter 5 and with the Markov model results. To do so, microscopic simulations have been run in this chapter according to the modeling assumptions used to develop the probabilistic models. To the author’s knowledge, this is the first comparison that considers non-stationary traffic process and that analyses the dynamic and stochastic behavior of queues also in the decreasing phase of congestion.

7.3. Set-up of VISSIM simulations

VISSIM is a microscopic software package, which is widely applied in research and in practical studies, for different evaluation and planning studies (i.e. signal control schemes, multi-class and multimodal networks etc.). VISSIM simulates traffic in a network based on individual vehicle characteristics representing multiple vehicle classes and heterogeneous driving behavior. The software users’ manual (PTV 2003) gives an overview of the possible applications of the software and the available built-in tools in VISSIM. This section briefly explains the basic methodology used in the program to simulate how flow propagates and generates queues in a network.

7.3.1. *Simulation assumptions*

Networks are modeled in VISSIM by choosing static characteristics (length and width of each road section or lane, connection between road sections or lanes, static road signs, detectors etc.) and dynamic features (traffic volumes and composition, route choice decisions etc.). The underlining driving criterion is that a vehicle tends to keep its desired speed all along the selected route, unless is constrained to modify this speed because of other vehicles on the road, or because of road signalization. The transition process from the desired speed to a new speed is thus modeled as a sequence of accelerations and decelerations towards the target speed.

Several parameters in VISSIM are defined as a probability distribution rather than a fixed value (drivers' desired speed, acceleration and braking skills, reaction times, etc.). The program user can feed the program with a specific model for the generation of the demand according to field observations or according to a known probability function. The loading process can vary also because of the stochastic flow composition. This option is not used in the comparative analysis presented in this chapter, since the impact of variable vehicle characteristics is not emphasized in this thesis.

It is logical that the distribution of queue lengths and delays strongly depends on the assumed arrivals, i.e. on the arrival headway distribution, since this represents the only source of variability assumed in the system. VISSIM models this distribution with a negative exponential function, which implies a Poisson distribution of vehicle counts in a discrete time period. This is consistent with the assumed distribution in the Markov and the Van Zuylen-Viti models.

Special attention is given in the program to the behavior of vehicles approaching an intersection. Drivers' level of alertness, threshold distance and lane change stimulus are different when vehicles approach a traffic signal (100m). The different attention level influences both average lane-changing and car-following behavior of vehicles. Other specific characteristics are considered in the modeling of a signal; for example reaction to amber light is modeled in the program by assuming a probability of a vehicle to stop if signal is amber. Other specific characteristics (i.e. different built-in control or routing decision methods, transit characteristics etc.) do not belong to the area of interest of this thesis and thus the reader may find more information in the manual.

7.3.2. *Representation of the network scenario*

To do microscopic simulations with the same assumptions as used in the previous chapters the network studied in VISSIM initially consists of only a one lane intersection with fixed control and homogenous traffic composition, namely passenger cars with all the same characteristics. The network consists of one lane road of 3 km in which a fixed control signal has been placed at 2.5 km from the origin. Two detection points were placed respectively before and immediately after the signal to detect arrival and departures at the chosen time step. Speed limit is set to 50 km/h for the whole road section, and the control is fixed, set with a cycle of 60 s and green of 24 s.

Although demand rate, traffic composition and flow distribution can be controlled since they are input variables, the capacity of the signal is not an input of the microscopic model but it can be determined directly from the results of the simulations. Moreover, signal settings are pre-defined in the program, but effective green and red times, used in the probabilistic models, have to be determined afterwards, since they will depend on the assumed traffic flow characteristics (acceleration, vehicle composition etc.). First step for the comparative analysis is therefore the determination of the signal capacity, together with the effective green and red time values. This is done by simulating the signal in a busy period, as described in the next section.

7.3.3. Determination of signal capacity and saturation flow

While the flow rate and its distribution are pre-determined in the program, the saturation flow is not pre-determined, as in most microsimulation programs. Since the capacity of the road is not an input parameter for the program but is the result of the roads, vehicles and drivers' characteristics assumed it needs to be estimated from a preliminary simulation.

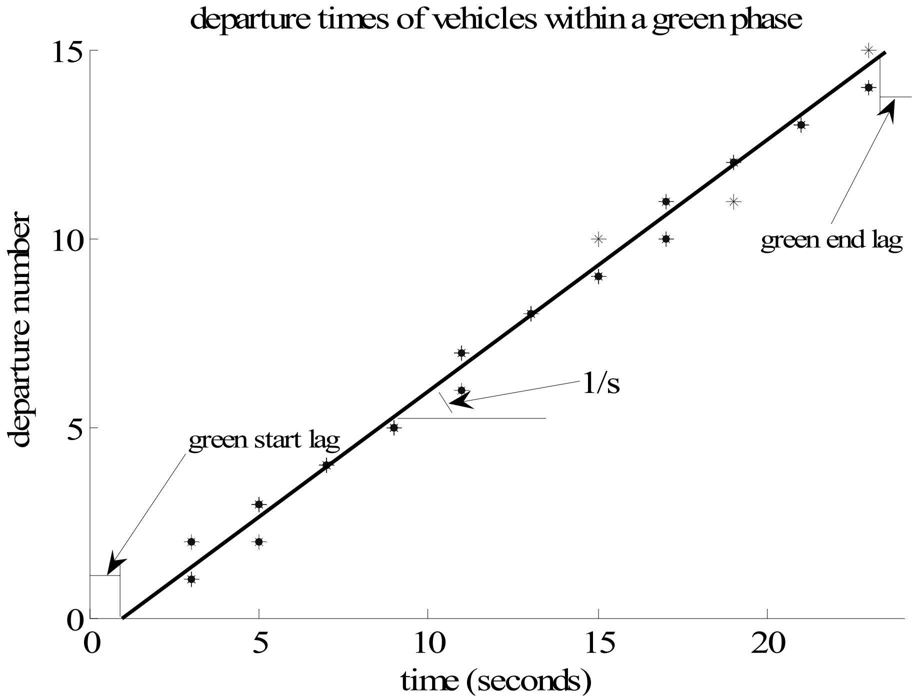

Figure 7.1 – Observed departure counts and departure time in a cycle and for several simulations

Using the regression method (Branston 1978), one can derive the average saturation flow of the road section and the effective green time. This criterion looks at the departure times of vehicles from the stop line and from the start of the green phase. The assumption is to consider in this computation only oversaturated cycles, i.e. an overflow queue occurs at the end of the green phase. A reason for this assumption is explained by looking at the histograms of vehicle counts shown in Figure 7.1 Effective green time and saturation flow are then determined by a regression method. Figure7.1 shows the simulated departure times of vehicles within a cycle, for 60 cycles and 10 simulations at a rate of 1000 vehicles per hour and only for saturated cycles. The detection points record the number of vehicles passing the road section within a pre-determined time interval (2 seconds). The regression line gives estimation of 1 second for an initial time lost for acceleration (*green start lag*) and 1 second of green end lag. The saturation flow is finally determined by its slope giving a value of approximately s=2200 vehicles per hour.

The effective green time is the green time from which the green start lag is subtracted and a green end lag is added, as said in chapter 3. The green start lag is due to vehicles accelerating during the initial part of the green phase. The green end lag is instead given by the behavior of people during the yellow phase, i.e. some people decelerate and stop while at the same time some others may not decelerate or even accelerate to pass the signal before it turns to red. Knowing the number of vehicles N that can pass during a cycle, one can check the correctness of these estimated by looking at the condition:

$$N = (t_g - \lambda_1 + \lambda_2) / s \tag{7.1}$$

where t_g is the green time as assigned to the traffic light, s is the estimated saturation flow (in vehicles per second), λ_1 is the green start lag and λ_2 is the green end lag. The term in parenthesis is an expression of the effective green phase.

Once the saturation flow and the effective green time are estimated from the microsimulations, one can determine the degrees of saturation corresponding to slight undersaturated and oversaturated conditions. The following simulations have been done for several different values of the demand near capacity, since the main characteristic of these two approaches is the modeling of stochastic overflow queues.

7.4. Overflow queue observations with stationary arrivals

First part of the comparison aims to analyze overflow queues when they gradually increase from an empty signal and for a very long evaluation period of stationary demand conditions having average below the capacity. Scope if the analysis is to validate the approximate expressions for the equilibrium of overflow queues in such conditions and the dynamic of these queues towards equilibrium. In oversaturated conditions the overflow queue does not reach any equilibrium but it increases as long as the road section is larger than the total queue length. The evolution of the expectation value is compared in this range with the deterministic queuing Formula (4.1), here reprinted:

$$Q_{t+1} = \max\{Q_t + a_t - d_t, 0\} \tag{7.2}$$

where a_t is the number of arrivals and d_t is the number of departures during $[t, t+1]$.

The computation of the overflow queue at the end of a time step can be easily made using the arrivals and the departures at each cycle and by computing Formula (7.2) for each simulation. Since vehicle interaction is low even in slight undersaturated conditions at sufficient distance from the signal and FIFO holds on the single lane road of the simulation, this assumption sounds reasonable.

7.4.1. Derivation of overflow queues in conditions near capacity

The microscopic simulation program has been run assuming various demand rates (i.e. degrees of saturation in the interval $[0.7,1.3]$). The demand is also assumed stationary for a total evaluation time of one hour. Green and cycle times are set in this section to respectively 24 and 60 seconds and the number of random seeds to 100. Tian (Tian 2002) consider a sufficient number of repetitions with less than half of this value (40) to obtain statistically valid results in terms of delay estimates.

Figure 7.2 shows the overflow queue computed from the observed arrivals and departures under low demand conditions (x=0.7), slight under-saturated (x=0.95) and oversaturated cases (x=1.1). The simulated queues are displayed by a sequence of dots. The continuous lines represent the resulting average values while the dashed lines are obtained by summing and subtracting the standard deviation to the mean. As one can see from the large spread of points, at undersaturated conditions also the microscopic program estimates a standard deviation approximately equal to the average, meaning that a driver may encounter, with the same average flow rates, no queue at all or the double of the expected value. These considerations confirm the conclusions given in Chapter 4 using the Markov model.

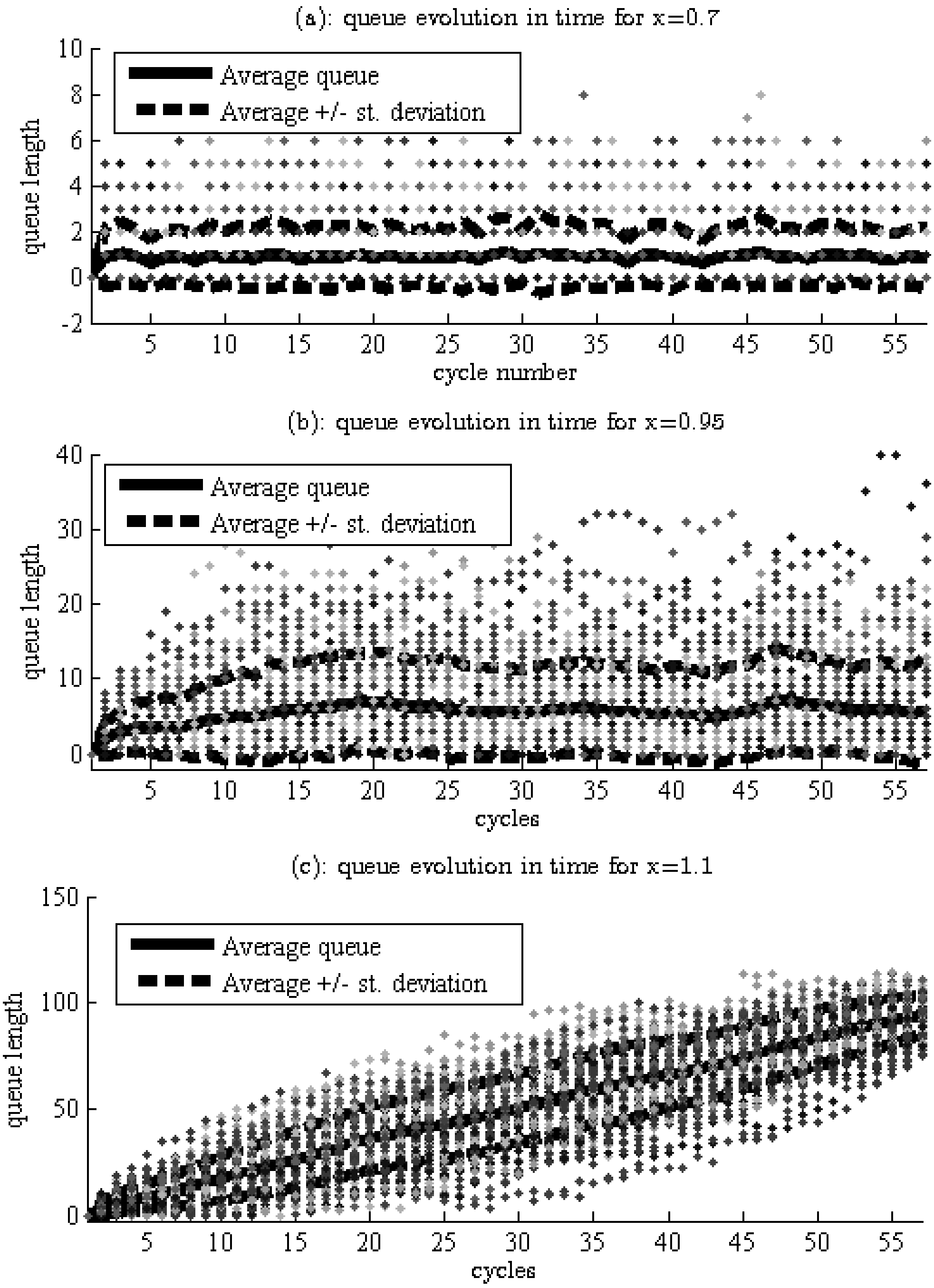

Figure 7.2: Overflow queue observations and resultant average and standard deviation

Both average and standard deviation show a concave form. A steady-state expectation value of the overflow queue is therefore observed in undersaturated conditions, while in oversaturated conditions this value follows a linear behavior. This can be explained by analyzing the sequences of overflow queues. In any simulated scenario the number of arrivals exceeds the number of departures within at least 15 cycles and overflow queue occurs 100% of the times thereafter. The resultant linear behavior of the expectation value is in accordance with Formula (7.2).

7.4.2. *Characterization of queue distributions*

Despite of arrival distributions, which are stationary for the whole evaluation period, the queue distribution of one cycle changes depending on its distribution at previous cycles, so no aggregation of different cycle times can be done resulting in a non-smooth behavior of expected values and standard deviations if one uses only 100 simulations.

The Law and Kelton method (Law 2000) allows one to compute the minimum number of replications in order to have statistically valid estimates. This method provides the minimum number of random simulations needed to obtain average arrivals, departures and queues within a *k%* interval of error with a pre-determined confidence interval α%. Since arrivals and departures do not vary in time, they can be aggregated in one distribution along the whole evaluation period. This means that fewer simulations are needed for the analysis of the arrivals and the saturation flow rate with respect to the analysis of queue lengths.

To compute the required number for analyzing the overflow queue length, firstly an initial number of replications in a pilot experiment is fixed, in this study 100. The computation is done by considering an average demand rate of 810 vehicles per hour (x=0.9) and at conditions near equilibrium (after 30 cycles). This choice is due to the higher variability of slight under-saturated conditions with respect to the oversaturated cases. It is under these conditions that one expects to need the largest number of replications. In this condition the queue assumes and average of 3.05 vehicles and a standard deviation *S* of 3.85. The Law and Kelton method computes the required number of replications with the following formula:

$$n^* = n \cdot \left(\frac{h}{h^*} \right)^2 \tag{7.3}$$

where

- n^* total number of replications required
- n number of replications of the pilot experiment
- h confidence interval of the pilot experiment
- h^* accepted confidence interval of the whole experiment

The confidence interval h, for the pilot experiment, is computed with the following formula:

$$h = t_{n-1;1-\alpha/2} \frac{S}{\sqrt{n}} \tag{7.4}$$

where $t_{n-1;1-\alpha/2}$ is the value of the *t*-student probability density function correspondent to a α% confidence interval (for 90% and *n*=100 is 1.66). With the fixed confidence interval one obtains a value *h*=0.64. The accepted error should be 0.1*3.05=0.305. This value is then not acceptable to consider the computed average to be correct. According to Formula (7.3) the minimum number of replications acceptable is 455. To get a sufficiently accurate distribution, and a smooth representation of mean and standard deviation, the number of simulations was increased to 1000.

The following analysis is referred to a demand of 810 vehicles per hour with a degree of saturation of 0.9. Figure 7.3 displays the histogram of observed queues.

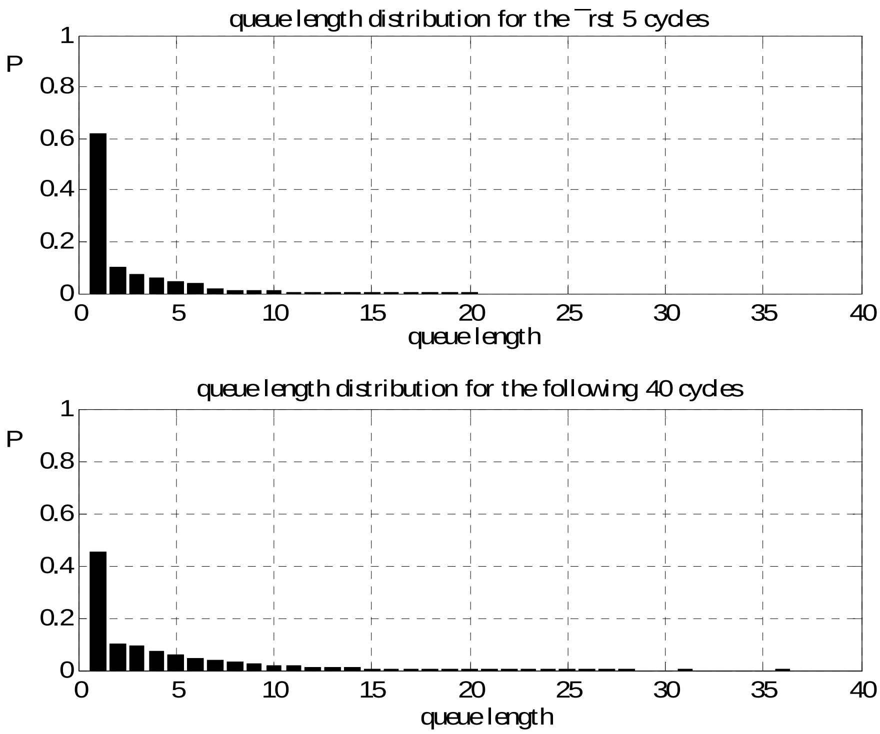

Figure 7.3: Distribution of overflow queue observations for a demand of 810 vehicles per hour

When the intersection is under-saturated the distribution is not clearly defined with a known probability density function. The queue length probability profile is in fact clearly influenced by the large probability of observing no overflow queue at the end of the cycle (almost 50% of the cases). The drop observed between the probability of observing zero vehicles and of observing non-zero vehicles justifies the decomposition of probabilities between zero and non-zero queues as described in the Markov models of Chapter 4.

This would be different if the simulation was run starting from a large overflow queue. The distribution of queues would be equal to the one displayed in Figure 7.3 (below) only after a sufficient number of cycles. In the transient phase, as long as the chance to discharge completely the queue is small, the queue shows a bell-shaped profile as in the oversaturated case, displayed in Figure 6.4 (second to fourth graphs). The program has been run in this case with an average arrival rate of 1000 vehicles per hour (degree of saturation of 1.1).

As it can be seen, in oversaturated conditions the distribution is initially influenced by the large chance of observing a zero residual queue, but after few cycles it assumes a more definable distribution, which can be approximated, within a cycle, to a Normal distribution. This is also in line with the conclusions given in Chapter 4.

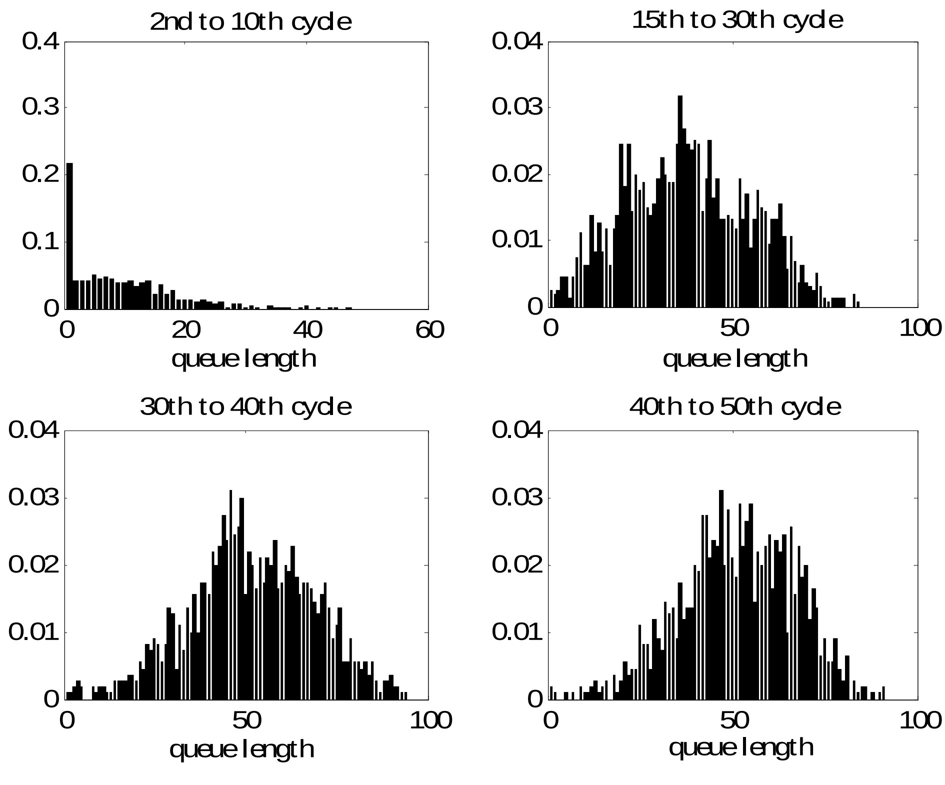

Figure 7.4: Distribution of overflow queue observations for a demand of 1000 vehicles per hour

The variability of queues is again shown to be very large and the distribution shifts gradually giving higher and higher average values.

7.4.3. *Equilibrium conditions*

In undersaturated conditions the queue distribution becomes stationary after few cycles; therefore, the expectation value reaches equilibrium. Looking again at the bottom picture of figure 7.3, there are observations where the queue is large, even over 35 vehicles, but still the probability of having zero overflow queue at the end of the green phase has a very large probability (here around 45%) while the probability that the queue is larger than 5 vehicles is relatively small (less than 20%). The expected value of the observed queues is consistent with the analytical expressions provided by Miller (Miller 1968) and Akcelik (Akcelik 1980), and the equilibrium values found with the Markov model. Therefore, the analytic expressions derived in steady state conditions sufficiently describe the behavior of the expectation value of the overflow queue when equilibrium is reached. The larger the flow rate, the longer time is needed to reach this equilibrium value from a non-equilibrium state.

Figure 7.5 shows the behavior of the overflow queue when an initial value of the queue of 20 vehicles is assumed.

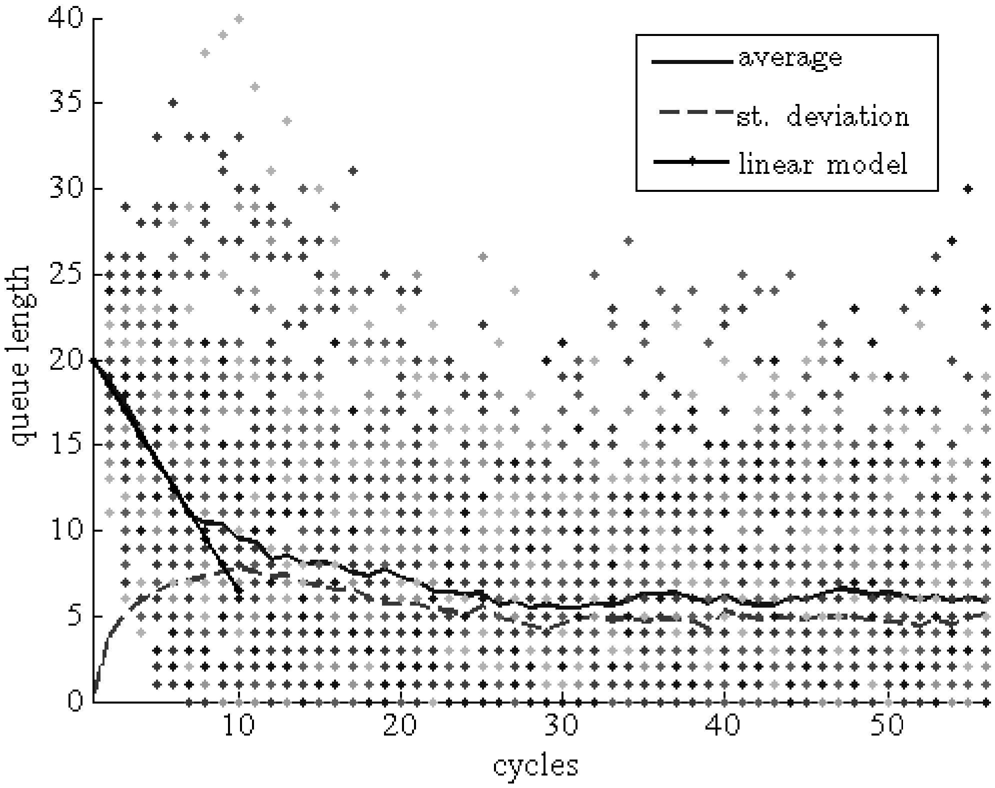

Figure 7.5: Overflow queue observations for a demand of 810 vehicles per hour and Q_0=20

The expected value of the overflow queue evolves with a linear behavior during the first cycles, when the residual queue cannot be served completely within one cycle. After some time the expected value does not evolve anymore according to Formula (7.2) but it asymptotically reaches the same equilibrium value reached assuming zero initial queues. The standard deviation has a peak at around 10 cycles. The chance to observe no queue at the end of the green phase starts being larger and consequently the standard deviation decreases in time also reaching the same equilibrium value computed when the initial queue is zero.

7.5. Comparison with macro- and mesoscopic results

This section compares the expectation value of the overflow queues and their standard deviation simulated with VISSIM and the ones computed with the Markov and the Van Zuylen-Viti models both in case of increasing and decreasing queues, with undersaturated and oversaturated conditions. To complete the analysis the overflow queue simulated in a peak hour scenario, shown in Chapters 4 and 5, is presented again using the microscopic approach.

7.5.1. Stationary demand conditions

The transition phase is analyzed by simulating stationary demand conditions for a sufficient number of cycles in the undersaturated case to observe expectation values close to equilibrium. Figure 7.6 shows again the average and the standard deviation of the queue for a demand of 810 vehicles (x=0.9) (top) and one of 1000 vehicles (x=1.1) (bottom), compared with the ones computed using the Markov chain process.

For the under-saturated case, the two characteristics simulated with the microscopic program are well reproduced with the Markov model. In both cases there is a slight error at the starting of around half a vehicle. At equilibrium the two curves have a statistically insignificant error. Given the consistency between the Markov model and the analytic formula, the consistency between the three approaches is clarified in these conditions since the Markov and the Van Zuylen-Viti models have been compared similarly in Chapter 6. Oversaturated cases are consistent as well, as Figure 7.6 (bottom) shows.

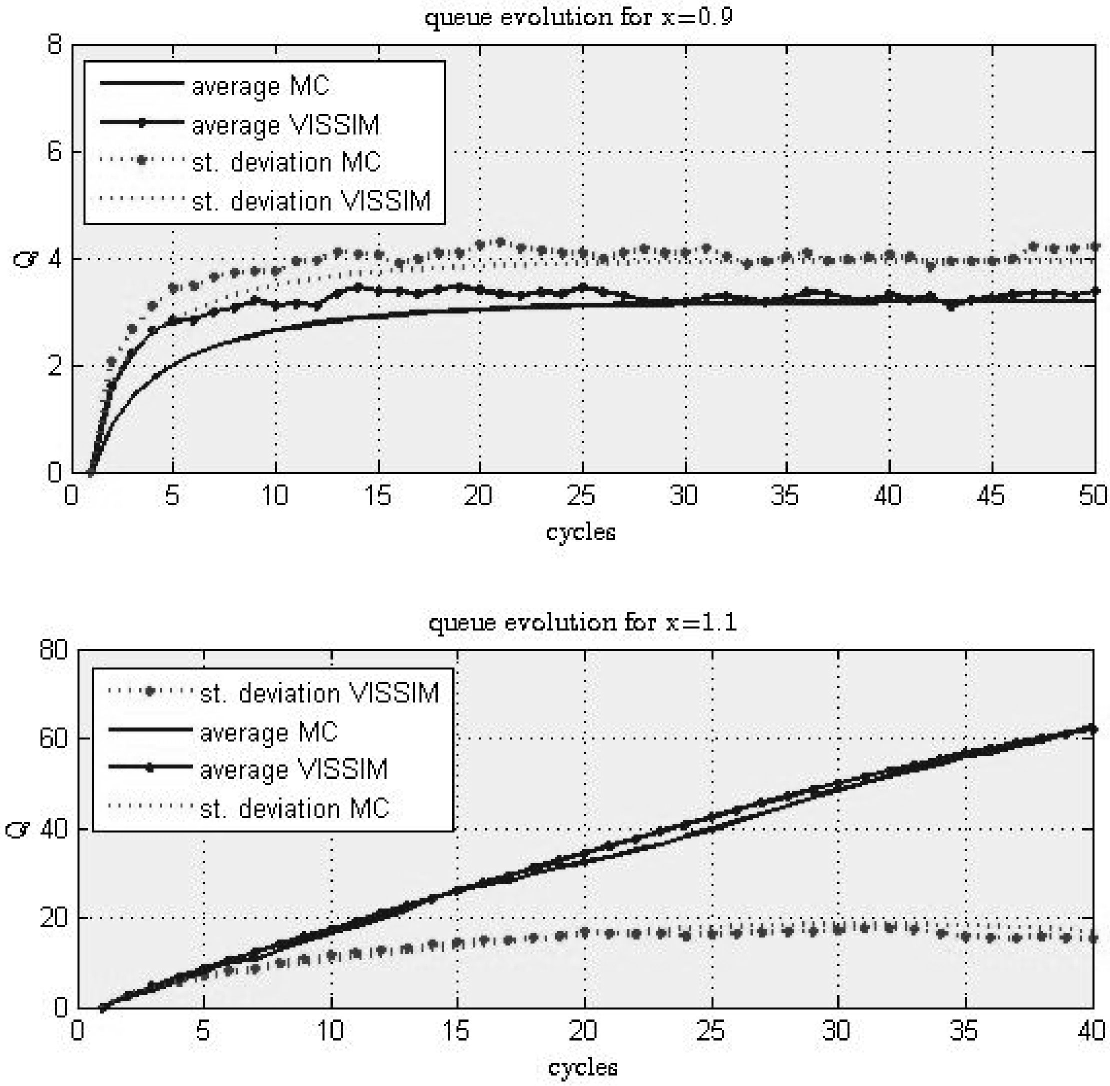

Figure 7.6: Comparison between microscopic simulation and mesoscopic simulation results for increasing overflow queues

The average and the standard deviation of the decreasing overflow queues illustrated as example in Figure 7.5 are compared in Figure 7.7. Although comparison of the expectation value with the mean value of the VISSIM results seems satisfactorily, large difference is found in comparing the standard deviation.

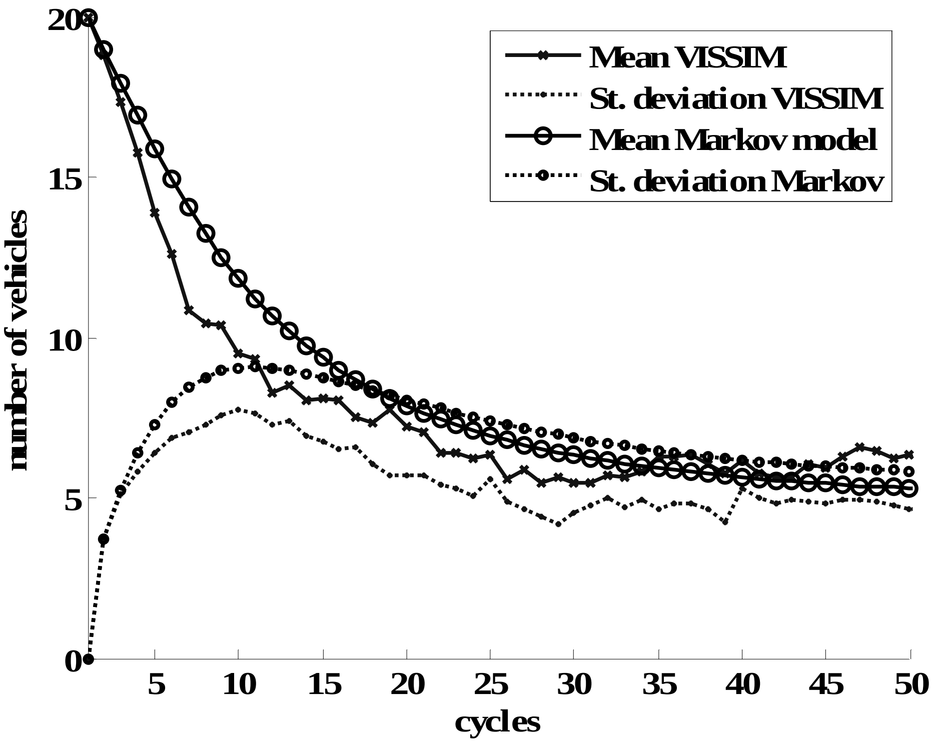

Figure 7.7: Comparison between microscopic simulation and mesoscopic simulation results for a decreasing overflow queue

In conclusion, the results from VISSIM show two main differences as compared with the Markov model:

1. When queues start from zero the expected value of the overflow queue increases more rapidly in VISSIM and for a considerable number of cycles in undersaturated conditions;
2. When queues start from large values the standard deviation of the queue increases less rapidly in VISSIM.

Both differences can be appointed at the arrival profile of VISSIM as compared with the assumption of Poisson arrivals in the Markov model. One possible explanation of the above inconsistencies between models can be in the car following characteristics of VISSIM. The arrivals are generated in the microscopic program at 2.5 km from the signal and only the loading of the vehicles is assumed as Poisson. In reality, while vehicles drive along the section they tend to disperse or form platoons according to the car-following behavior assumed in VISSIM. Therefore, the arrival at the signals should not be Poisson. This effect is quite visible at large demand conditions. This explanation is confirmed by Tian (Tian 2002), who found large differences in the results of three microscopic simulation programs (CORSIM (FHWA 1999), VISSIM (PTV 2003) and

SimTraffic (Corporation 1999)), especially in terms of variability of the performance measures. He concluded that "it is suspected that when a link is highly saturated, drivers tend to behave more uniformly, resulting in reduced variations on the performance measures". Akcelik (Akcelik 2001) agrees upon this dynamic character of the saturation flow rate. He recommended analyzing this behavior in a microscopic program in function of the queuing, acceleration and car following model parameters of the microscopic program.

7.5.2. Variable demand conditions

To conclude the comparative analysis, a peak period is simulated by assuming a step-wise average demand. Demand is then assumed stationary only for a limited period of time, which in this section is assumed 10 minutes. Figure 7.8 shows the results of the simulation from the microscopic approach.

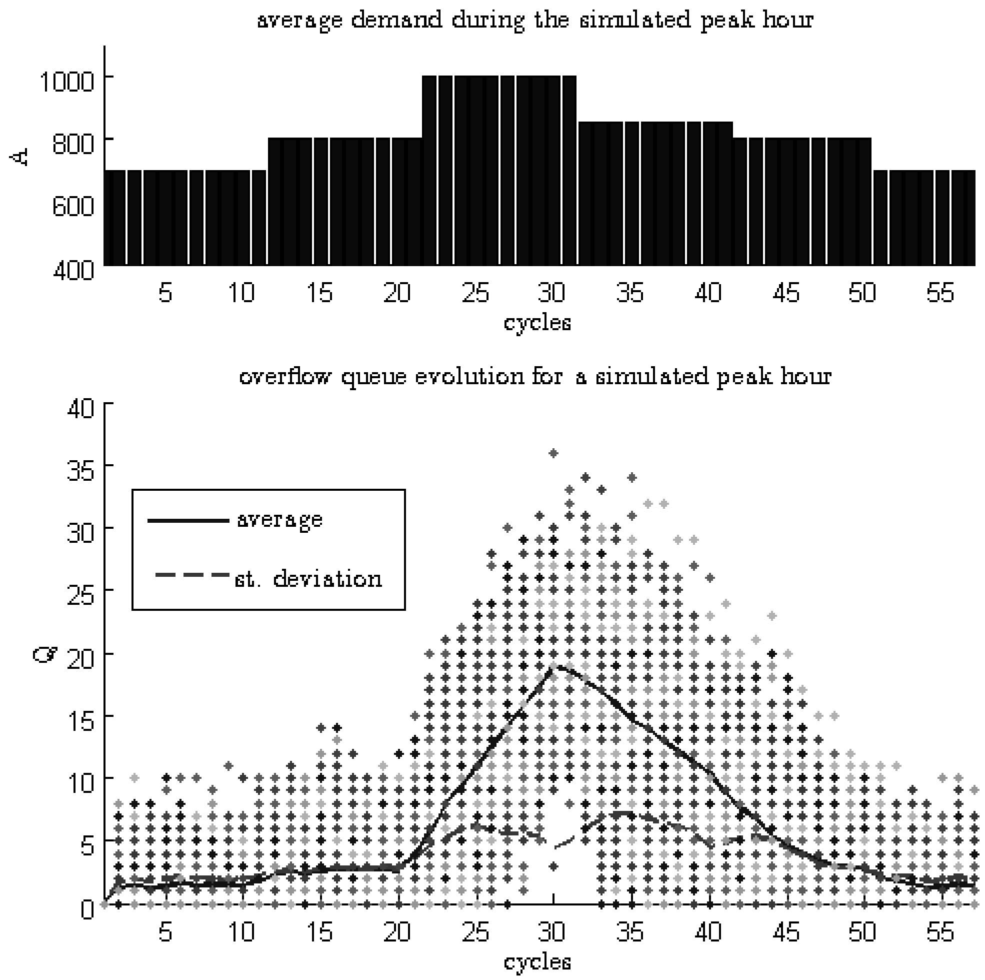

Figure 7.8: Simulation of a peak hour in VISSIM

The top picture shows the assigned demand while the bottom picture displays the simulation results in terms of queue lengths for at each simulation, together with average and standard deviation of all simulations. In terms of dynamics of the average and standard deviation of the queue, the exponential behavior at the shoulders and the linear behavior at the central periods it is observable, according also to the results presented in (Viti 2004).

The program has been run for this case with 200 different random seeds. One can clearly see the very similar behavior of expected value and standard deviation in conditions of undersaturated traffic, while the variance-to-mean ratio decreases consistently in slight oversaturated conditions, reaching around 0.25. When the queue starts decreasing the behavior follows the linear decrease as expected while for the last two sub-periods the queue evolves exponentially.

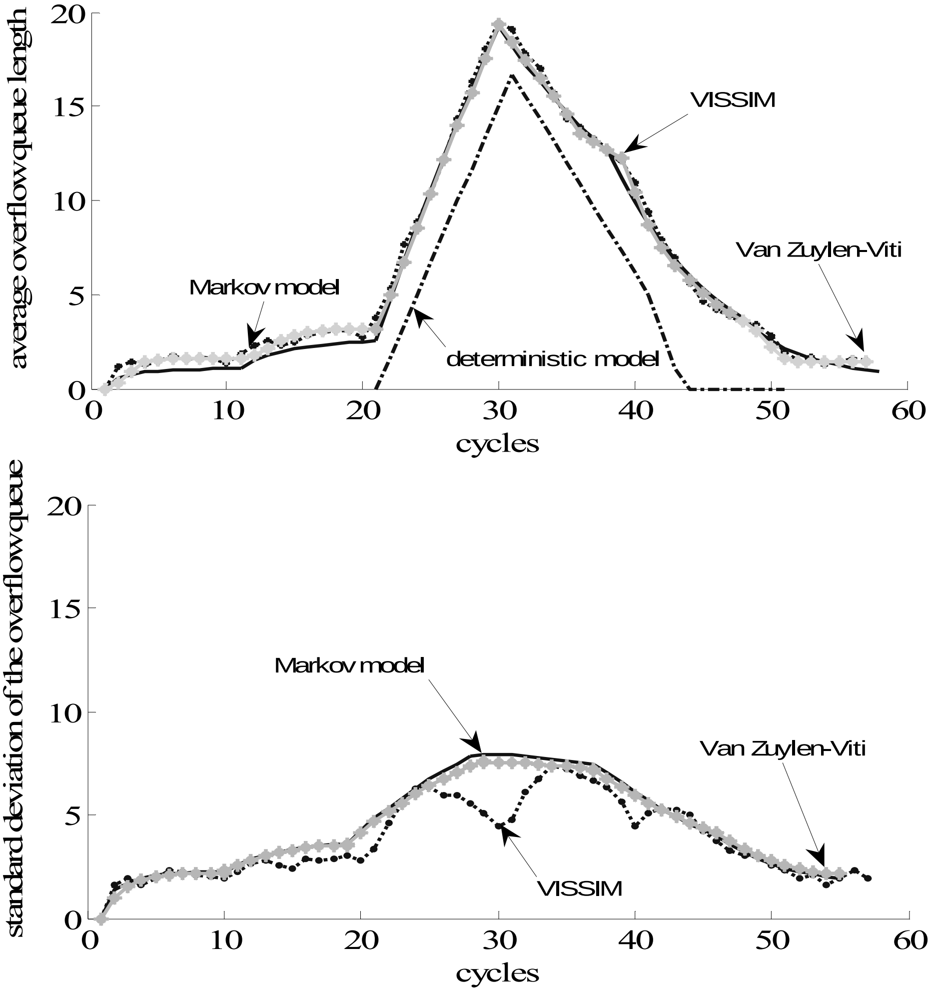

Figure 7.9: Comparison of overflow queues between Markov, Van Zuylen-Viti and linear models with the results of VISSIM

Figure 7.9 compares the results in terms of average (Figure 7.9 (a)) and standard deviation (Figure 7.9 (b)) measured with VISSIM, with the Markov the Van Zuylen-Viti models and the deterministic formula adopted in the Highway Capacity Manual (TRB 2000) as done in the example in Chapter 6.

Comparison between expected values derived from the software program VISSIM, the Markov model and the Van Zuylen-Viti formula are very good (Figure 7.9, top), and the error remains smaller than 2 vehicles for the whole evaluation period, while the error made using the deterministic Formula (7.2), which is used in the HCM, appears consistently larger. This represents an important contribution to traffic managers and practitioners, since it proves that the dynamics of the overflow queue are well estimated with all three different level-of-detail models under non-stationary demand assumptions.

Moreover, the assumption of Poisson arrivals does not seem to be limiting in terms of average overflow queue length, since the comparison between microscopic and mesoscopic simulations shows good consistency even if the Markov model does not consider any car-following behavior.

The consistency between models showed by comparing the average values is not completely confirmed with the comparison of standard deviations. The Markov and analytic models show a smoother behavior but also considerably higher values especially at the highest peak. This difference can be again appointed at the non-uniform arrival distribution of vehicles during peak periods due to the car-following behavior assumed by the microscopic program.

7.6. Overflow queue variability in arterial corridors

The arrival profile should differ if another signal is placed in a short distance upstream, as discussed in Chapter 3. This problem has been simulated in VISSIM assuming a sequence of two signals.

Effect on the overflow queue behavior is given in this system by the filtering effect. The reduced number of possible arrivals within a cycle results in a lower chance to observe an overflow demand in roads where the secondary roads play little role. According to Newell (Newell 1971), overflow queues in those corridors would appear mostly at the first signal, if the capacity of the downstream signal is at least equal to the upstream one. The upstream signal works in these conditions as filter for the downstream one. Since a limited number of vehicles can pass the signal within the green time, all exceeding vehicles would wait at the upstream signal till their service time. The additional number of vehicles at the downstream signal is thus very small. As consequence, there is little chance to observe long overflow queues at downstream signals, even if the demand entering the network at each cycle exceeds the capacity. To observe this phenomenon using microsimulation and give an estimate of the reduction of overflow queues at downstream signals, two signals were placed at relatively short distance.

Figure 7.10 displays the detected vehicles at the origin of the road section (figure 7.10-(a)), right after the first signal (Figure 7.10-(b)) and immediately after the second signal (Figure 7.10-(c)) for a demand of 810 vehicles per hour ($x = 0.90$). Signals were placed at 500m distance. It looks that the arrival and departure distributions at the second signal are very similar one to another, suggesting that nearly all vehicles leaving the first signal are served at the second signal within one cycle.

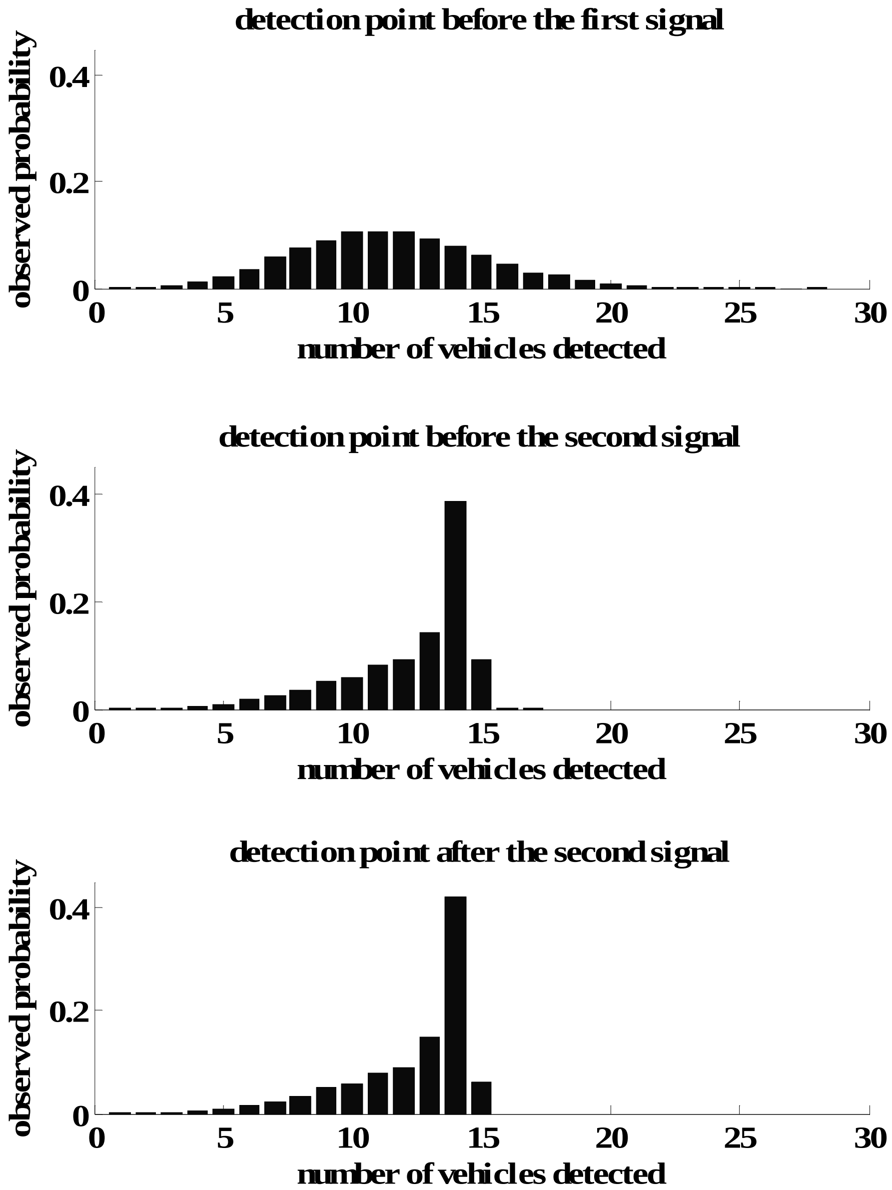

Figure 7.10: Vehicles detected per cycle at the three detection points for a demand of 810veh/hr

Looking at the evolution of overflow queues in figure 7.11 it seems that the increase of the average demand rate arriving at the first signal from 810veh/hr to 1000veh/hr (x=0.9 and x=1.1 respectively) does not influence significantly the distribution of queues at the second signal, as one can see by looking at figures 7.11-(c) and 7.11-(d). In both cases the average number of vehicles, which statistically is not served within one cycle, is nearly 1. In conclusion, overflow queues are very small at downstream intersections and their contribution to the overall delay of vehicles is expected to be very small too. Obviously this effect reduces if the capacity of the downstream signal is reduced, or if other flow streams converge to the downstream intersection.

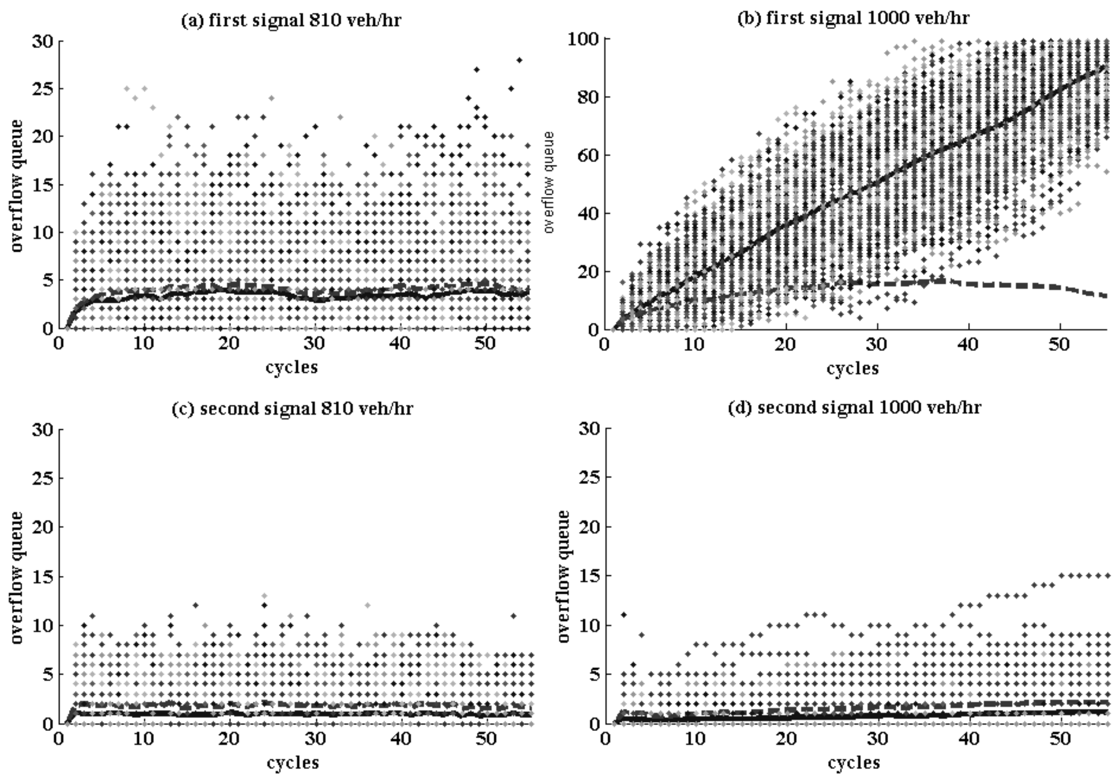

Figure 7.11: Overflow queues in paired intersections for demands of 810veh/hr and 1000veh/hr

In conclusion, the estimation of delays incurred by overflow queues in arterial corridors is in accordance with the findings of Newell. This conclusion applies only if there is very little contribution from turning flows upstream and the capacity of the downstream signal is at least equal to the one upstream. With these settings, the isolated intersection represents the worst case scenario one can compute for the estimation of overflow delays at traffic signals, while the above represents the most optimistic scenario for the downstream intersection. The next chapter extends the Markov model for isolated intersections also to arterial corridors by simply adding a constraint to the maximum number of arrivals within a cycle randomly generated by the model depending on the flow rate arriving from the upstream section.

7.7. Summary

This chapter compares the mesoscopic method based on Markov chain processes described in Chapter 4 and the analytic formulas developed in Chapter 6 for the expected value and the standard deviation of the overflow queue length with several simulation runs of a commercial software package based on microscopic programming, VISSIM.

Microsimulation is the only practical alternative to field data for this study, since it is rather unlikely that one can observe in real life sufficiently long periods of stationary demand conditions. To operate the analysis presented in this chapter (i.e. behavior towards equilibrium, behavior with an initial queue, behavior with non-stationary demand rates etc.) data were needed which repeated the overflow queuing process under the same assumptions.

The stochastic processes make it necessary to repeat several microscopic simulations for situations that are close to saturation. Hundreds or even thousands of simulations are needed if an accurate estimate of random queues or delays is required. This justifies a lack of a thorough analysis, which is done in this chapter. To the author's knowledge, this is the first study that attempts a comparison of models in a non-stationary demand rate scenario that gives special importance to the behavior of overflow queue lengths from large initial values. Only a few studies have instead compared models in the way they deal with the variability of traffic at signals.

The Markov and the Van Zuylen-Viti formulas have been compared with the results of the VISSIM microsimulation showing very good agreement. The consistency between the three approaches in various conditions of traffic validates the two less detailed methods. This represents also an important contribution to traffic managers and practitioners, since it proves that the dynamics of the overflow queue are well estimated with all three different level-of-detail models. The analytic function presented in chapter 6 is therefore suitable for planning and design purposes and contributes to a better estimate and prediction of the signalized network performances. The consistency between the models also in a dynamic scenario with non-stationary demand rates implies that the Van Zuylen-Viti model may contribute to the development of improved network loading models in of the models has also shown an inconsistency between the models considered in the initial behavior of expectation values and standard deviations. The standard deviation shows consistently larger values in the Markov model with respect to the simulated results of VISSIM especially when recovering from a large initial overflow queue at the start of a new sub-period. This has been explained by a more uniform behavior of the vehicles in the microscopic program with the increase of congestion due to the assumed car-following logic.

Finally, microscopic simulations in an arterial corridor show that overflow queues at downstream signals give little contribution to the overflow delay component if signals have the same settings. The upstream signal works in these systems as filter, reducing the chance of cycle overflow to the downstream ones. In this sense, isolated intersections represent the upper bound. A reduction factor should be applied (for example like the model of Van As, (Van As 1991) presented in Chapter 3) in order to compute the overflow queue at downstream signals. Several widely applied models (e.g. TRANSYT, (FHWA 1984)) do not account for this effect, resulting in an overestimation of overflow queues at downstream sections. The next chapter considers the application of the Markov model for more complex scenarios than the isolated, fixed time, single lane sections, among which the effect of upstream signals is discussed.

8

Probabilistic delay models for arterial corridors, multiple lanes and dynamic controls

8.1 Introduction

Chapter 7 showed that a probabilistic model is able to give statistically consistent results with respect to microscopic simulation at isolated signalized intersections. Both mesoscopic model and microscopic simulations have been computed under very simple assumptions. The method presented in Chapter 4 has been in fact developed in the assumptions of traffic uniformly distributed within a cycle time (isolated intersection), fixed-time control and FIFO (First In, First Out) discipline, which is a valid assumption only at single lane sections or sections with prohibited overtaking. Nevertheless, these assumptions characterize most of the analytic models used in practice, as Chapter 3 showed. In Chapter 6 the results of the Markov model developed in Chapter 4 have been used to develop models for the expectation value and the standard deviation of the queue under weaker assumptions than the currently adopted models.

The probabilistic approach gives the traffic analysts the opportunity to include dynamic and stochastic features to the travel time estimation problem. The potentials of this method have not been highlighted yet, and its application is still quite limited in practice by its simplifying assumptions. Fixed controls are being increasingly replaced by more advanced control devices, as already discussed in Chapter 3. Moreover, the assumption of isolated signal represents an upper bound for the estimation of overflow queues in urban networks, especially when signals are closely spaced, as shown in the case of two intersections simulated in Section 7.6. A question is also whether FIFO condition holds at such intersections that may have more lanes dedicated to a traffic stream. The lane selection criterion at multiple service points represents an interesting research question also for other purposes, for example for the design of road sections and service points at toll plazas, or if one takes into account spillback effects in the estimation of delays and in the design of exclusive turning lanes.

This chapter addresses these issues by showing the possibility to model the distribution of queues and delays using the Markov model also in the context of:

- *Arterial corridors*: stochastic delays are influenced by the filtering effect of signals, while they are not primarily affected by the platooning effect, which instead influences the uniform delay component, as explained in the literature research presented in Chapter 3. Filtering effects are considered in the Markov model by simply adding a constraint in the distribution of arrivals at the intersection, as it will be shown in Section 8.2.

- *Multilane sections*: stochastic delays can be influenced by an unbalanced distribution of traffic among lanes. Different streams merging or splitting at any interrupted flow section may not distribute uniformly among all lanes. Unbalanced distribution of traffic among lanes affects the lane changing behavior of travelers, who might decide to try and move to the lane with the smallest queue in order to minimize their expected waiting time. Lane changing possibilities are on the other hand limited by the density of traffic on the target lane. Section 8.3 describes how the Markov model can be extended to multilane sections.

- *Dynamic controls*: the Markov model can be extended to estimate queues and delays in demand-responsive controls, characterized also by variable signal settings. This modeling issue is analyzed in Section 8.4 where a probabilistic model for vehicle actuated controls is developed.

The extensions proposed in this chapter contribute to the modeling of delays, but they contributes also to the analysis of the variability of travel times in general urban networks, given the possibility to compute the distribution in time of such delays. These are only examples of the modeling opportunity of the probabilistic approach. It is envisaged by the author that this modeling technique can be used in any (traffic) process that is characterized by a strong dynamic and stochastic behavior.

8.2 Overflow queues in arterial corridors

The effect of upstream signals can be subdivided into two elements: the platooning and the filtering effects. The first effect is due to the bunching of vehicles during the red phase and the movement of vehicles in platoons during the green phase. The second is due to the limited number of vehicles, which can pass the upstream intersection within the green phase and it is primarily determined by the signal capacity. This effect has been considered in the delay models by simply assuming a progression adjustment factor in the uniform delay component that accounts for the signal coordination quality (Chap. 3).

Even accounting for the natural dispersion effect of vehicles due to the car-following logic the platooning effect has negligible impact on the overflow delay component. For this reason platoon and dispersion effects are not considered in the following considerations regarding the overflow queue and the development of a Markov model extension to account for a different arrival profile due to the effect of upstream signals.

The filtering effect modifies the arrival distribution at the signals and therefore the stochastic and overflow delays. The distribution of arrivals at isolated intersections is characterized by an upper bound, which is at least determined by the saturation flow. Therefore, a positive number of vehicles can be observed up to a maximum value. An intersection inside an urban network is connected to other intersections and the maximum number of vehicles, which can be observed within a cycle, depends on the way other controls operate, and on the way they are connected one to another.

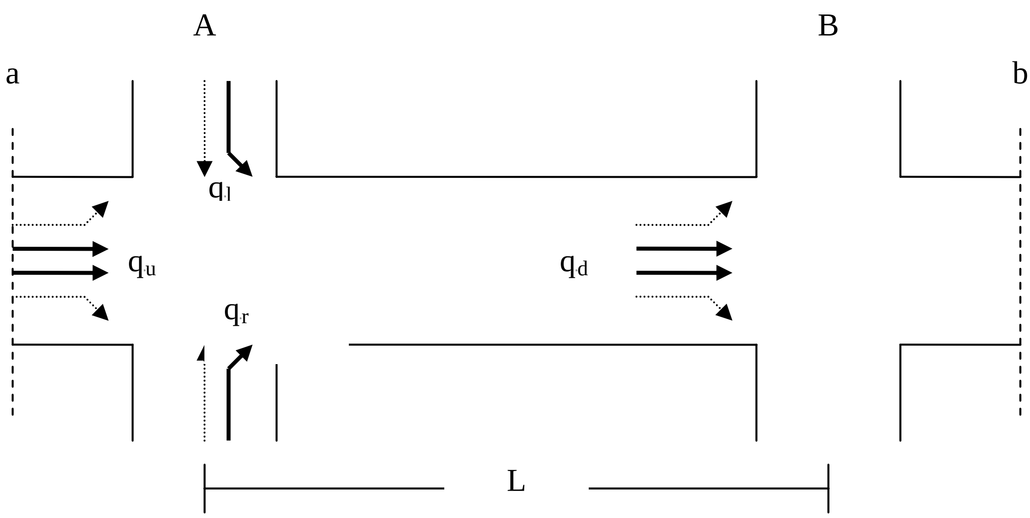

Figure 8.1: Scheme of two intersections in an arterial corridor

Figure 8.1 shows for example a schematic representation of two intersections shortly distanced each other. For sake of illustration the road section in between sections **a** and **b** is a one-way road connected to two secondary roads in the nodes denoted by **A** and **B**. The arrival profile q_d at node **B** is determined by all streams converging from node **A** and directed towards the second intersection, thus the flows denoted with q_u for the arrivals at the upstream intersection, q_l and q_r respectively for the flows converging from the left and from the right secondary roads to the arterial road.

The estimation of the arrival profile at the downstream intersection is strongly influenced by the green time given to all three streams ensuing from the upstream signal and the volume of traffic observed on these streams. The maximum number of arrivals in the distribution of q_d has still an upper bound, determined by the saturation flow of the upstream section. Let s be the maximum number of vehicles within a cycle, which is determined by the saturation flow. In the assumption of no shared green time among traffic streams, the upper limit of the distribution of arrivals in a cycle at the downstream intersection can be mathematically formulated by the following condition:

$$\max\{q_d\} = \min\left\{\sum_i \left(\frac{t_{g,i}}{t_{C,u}} \cdot s_i\right), s_d\right\} \tag{8.1}$$

where s_d is the saturation flow at the downstream section, while g_i and s_i are respectively the saturation flow and the green time assumed from the upstream flow coming from direction i and C_u is the cycle time of the upstream signal.

In Chapter 7, Figure 7.10 the filtering effect was shown in terms of overflow queue using microsimulation. It turned out that in a signalized corridor the effect of an upstream signal strongly reduces the overflow queue computed at the downstream signal, and that the arrival flow profile is strongly affected by the upstream signal. This effect can be reproduced using the Markov model by simply assuming an upper bound to the maximum number of arrivals.

The result on the queue and delay dynamics is that the maximum outflow of the upstream intersection has a large influence on the development of the overflow queue and its standard deviation. Figure 8.2 confirms this result. A Markov model has been calculated for a volume-to-capacity of 0.95 and assuming different maximum arrivals If the maximum arrival value of the upstream intersection is high, small influence is found from the network, as one can see from the small difference between the dynamics of the expectation of the overflow queue assuming a maximum outflow of 21 and 23 vehicles. If the maximum number of arrivals is limited by an upstream intersection, the variation in the number of arrivals is limited too. If the maximum outflow of the upstream intersection is smaller than the capacity at the downstream signal no overflow queue will be observed.

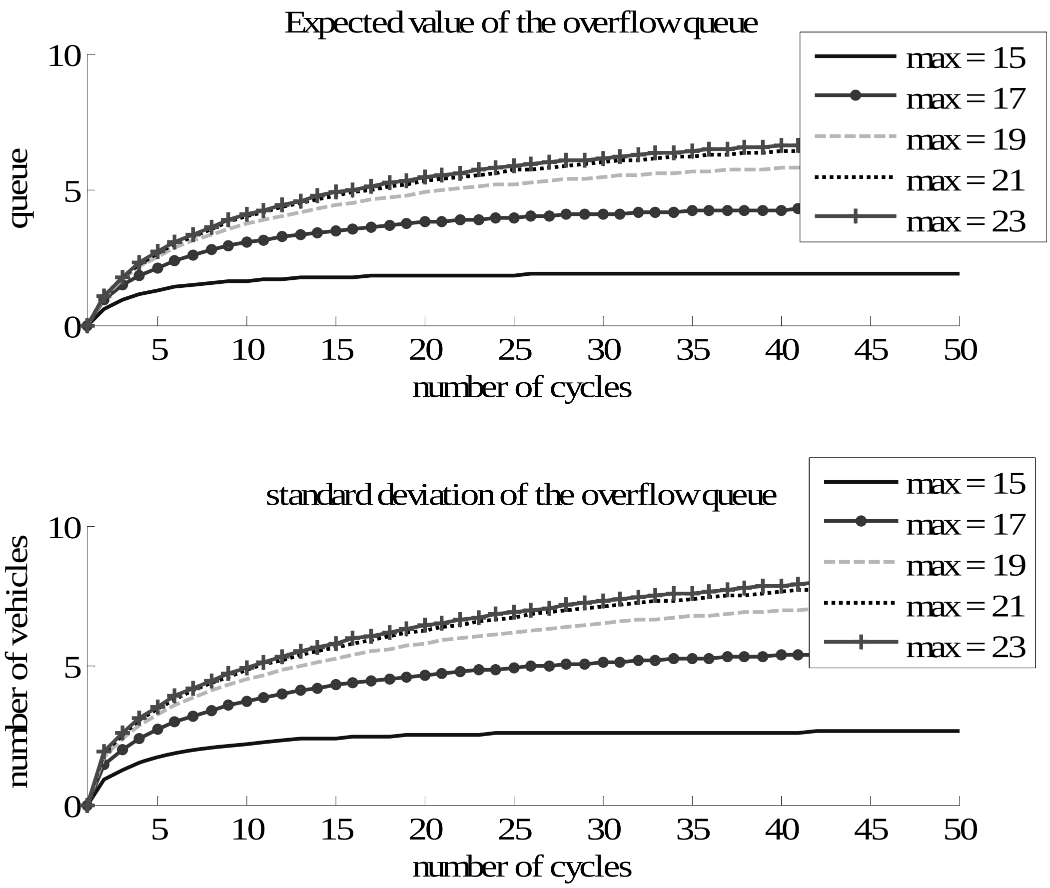

Figure 8.2: The dynamics of overflow queues assuming different maximum arrival rates

In conclusion the Markov model can be applied in the context of arterial corridors by simply assuming a maximum number of vehicles in the arrival distribution if the traffic comes predominantly from one main stream. A different distribution should be thought if the traffic comes from two or more streams (e.g. bi-modal distribution for two main streams). Historical traffic observations can suggest a realistic arrival profile to implement in the Markov model at each intersection of a network. Formula (8.1) can still give an estimate of the maximum arrival rate that can be observed.

8.3 Multilane intersections

8.3.1 Problem description

The Markov models presented so far in this thesis assume single service point. Within this system the drivers have no option than following their preceding vehicles and FIFO condition holds in the arrival and service processes. It is rather frequent in practice to design intersections with more than one lane dedicated to a flow stream, as Figure 8.3 shows. Furthermore, other systems where the queuing model can be applied are characterized by multiple service mechanisms, e.g. toll plazas on motorways. Unequal distribution of flows among lanes of a road section may be frequently observed in practice. For example, if the left lane is dedicated to overtaking operations, usually a larger percentage of vehicles will be observed on the right lane. The degree of lane usage can be also affected by the road geometry of downstream sections; for example a certain number of vehicles may decide to drive on one lane to facilitate turning operations at later sections.

The distribution of flows may affect the queue and delay at signals and vice versa. In fact, if the road geometry allows the drivers to check the distribution of queues at the downstream intersection, they may try and move to the lane with the shortest queue observed at the time of their decision.

Figure 8.3 shows an example of multilane section before a signal where two lanes are dedicated to the straight-through stream and one lane to right-turning vehicles. If a larger percentage of vehicles drive on the right lane, it can happen that the queue building up on the right straight-through lane blocks the access to the exclusive right-turning lane, reducing the capacity and the operational efficiency of the intersection. In order to avoid this problem, some vehicles may try and move to the left straight-through lane.

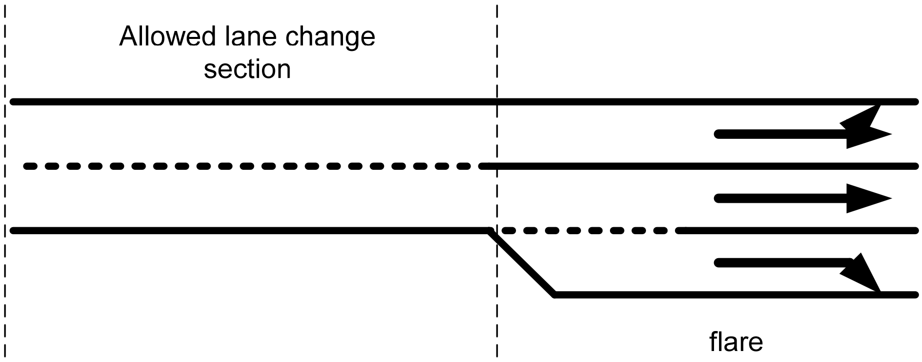

Figure 8.3: example of a multilane section before a signal

Often in practice exclusive turning lanes (or flares) are too short due to design constraints or bad design. An underestimation of a flare can lead to several problems, i.e. spillback with consequent increase of waiting times and lane blockage, which can imply serious decrease of safety. Cars may not be served within a green phase because they are blocked by cars spilling back from another lane; moreover they can try risky maneuvers to be served, increasing the chance of accidents.

Not all vehicles that are willing to change lane have actually the possibility to do it; lane changing possibilities are, on the other hand, limited by the gap acceptance of the users.

A user might be willing to change lane but this intention can be somewhat limited by the presence of other vehicles on the target lane at the moment the traveler should change lane. This operation can be observed also if the queue building up on the exclusive turning lane spills back. In fact, if a queue builds up on the exclusive right-turn lane, and exceeds the length of this lane, it will block the right straight-through lane as well, producing hindrance to the vehicles arriving on that lane. This phenomenon will push the flows to increase the lane-changing maneuvers towards the left lane, reducing the total intersection capacity.

The described lane changing phenomenon involves several different operational and psychological factors. Microscopic simulation models can give realistic estimates of this variability since they simulate the traffic propagation at the vehicle level. On the other hand also microsimulation models simulate the arrival distribution of vehicles using some lane selection logic (see Chapter 3). Appendix C gives an example of results from a macroscopic viewpoint from several microsimulations in a two-lane scenario. These simulations confirm that vehicles have a strong lane changing behavior before a signal, and this behavior changes with the length of the queues waiting at the signal.

8.3.2 Probabilistic model of lane changing behavior

The unequal distribution of flows may affect the queue and delay at signals and vice versa. In fact, if the road geometry allows the drivers to look ahead to the distribution of queues at the downstream intersection, they will try and move to the lane with the shortest queue. The possibility to change lane is, on the other hand, dependent on the headway distribution of vehicles driving on the target lane and the equality may not often occur. Variability of these flows can be therefore observed. Spillback effects from other lane groups can also influence this lane changing behavior. This effect is a clear example of lane underutilization that should be taken into account in the design of the approach and in the computation of the effective capacity of the signal (Tian 2006).

The queuing process depends on the randomness of arrivals and departures, but the distribution of the demand can be influenced by the queue distribution itself because of lane-changing behavior. It is assumed here that the user increases his “intention” to change lane the larger the difference between the queue at the driving lane and the target one. Since the queue length is characterized by a probability distribution, this intention can be described by a probability distribution too. Lane change is, on the other hand, limited by the degree of occupancy on the target lane, since vehicles need a sufficient gap to operate the maneuver. One can apply a Bayesian updating rule to compute this chance, thus compute the conditional probability that a vehicle changes his current lane. This probability is simply computed by the product of the two probabilities, since it is assumed that they are independently distributed stochastic variables.

Let a denote a lane of a road section and b be an adjacent lane. Let suppose that the user has time to evaluate the queue length distribution among lanes at the downstream signal and (eventually) change lane. Let suppose also known average number of vehicles arriving at the intersection for each lane group, q at a certain section upstream. For example the flow fraction on lane a at time t will be denoted with $\alpha_a(t)$. Knowing the average and distribution of the split rates one can compute the flow distribution $q_a(t)$ for lane in a cycle by using the relationship $q_a(t) = \alpha_a(t) \cdot q$. This component is also a random variable since this fraction can vary from cycle to cycle. Let suppose to know the queue distribution at the starting of the simulation for each lane, $q_a(0)$. The travelers' probability of intention to change from lane *a* to lane *b*, $\Psi_{a \to b}(t)$, is assumed to depend on the difference in queue length $Q_a(t) - Q_b(t)$ on lane *a* and *b* as given by the following formula:

$$\begin{cases} \Psi_{a \to b}(t) = h(Q_a(t) - Q_b(t)) & \text{if } Q_a(t) - Q_b(t) > 0 \\ \Psi_{a \to b}(t) = 0 & \text{otherwise} \end{cases} \tag{8.2}$$

The above probability is supposed to be a known function h, which increases and gets closer to 1 the larger the difference between queues. Inversely, if the queue at the traversing lane is smaller, there is no reason to consider a possible lane change to the adjacent lane[1].

A user can change lane only if there is enough gap for the maneuver. Therefore, known the number of arrivals, one can deduce the headway distribution of cars and consequently the probability of having enough space to change lane. For example, if the arrivals are Poisson distributed, the time headway distribution can be approximated as negative exponential. The chance of having a sufficient gap to change lane is then equal to the probability $\Pr_b(l \geq \bar{l})$ that the headway between observed cars, l, is higher than a predefined threshold, $\bar{l}$. Once these probabilities are computed, the number of vehicles $\eta_{i \to j}$ that move from lane a to lane b is given by the following formula:

$$\eta_{a \to b}(t) = q_a(t-1) \cdot \Psi_{a \to b}(t) \cdot P_b(l \geq \bar{l}) \tag{8.3}$$

The total number of arrivals q_a at lane a will then be given by:

$$q_a(t) = q_a(t-1) - \eta_{a \to b}(t) + \eta_{b \to a}(t) \tag{8.4}$$

[1] This value may be positive if one considers the case of vehicles moving to the left lane to guarantee accessibility to the exclusive turning lane. This issue is not considered in this study.

This value is then used to compute the distribution of the arrivals $\Pr_q(k)$.

It should be noted that, in the one-lane problem, the distribution of arrivals is independent on the present queue, and the transition matrix is invariant with time, while in this case it becomes queue-dependent and thus time-dependent.

8.3.3 Effect of a short flare

Suppose now that there is an exclusive turning lane, c, at the intersection and to know the flow rate, q_c. For sake of simplicity, one can assume that the vehicles that arrive at the intersection and have to turn, are already at the closest lane before entering in the exclusive lane, thus no intermediate lane changing to reach the exclusive turning lane is considered. As it happens often in real conditions, one can assume the green time for this exclusive lane to be different from the straight through lanes, g_c. Under these assumptions, the queue at the exclusive turning lane is computed exactly as a single lane intersection using the standard single-lane Markov Chain.

Let α_c be then the fraction of the total demand q representing turning flows and let $Q_{spillback}$ be the maximum number of vehicles, which can be placed in the exclusive turning lane without creating spillback. Due to the randomness of arrivals, there is a non-zero chance of having spillback. The probability that spillback occurs can be computed with the Markov model. In this condition the adjacent lane will be influenced by this phenomenon. If the length of the adjacent lane queue is smaller than the accumulation lane queue the user will consider the latter for his lane changing behavior. Thus, all equations above do not change apart conditions (8.2) that become:

$$\begin{cases} \Psi_{a\to b}(t) = h(\max(Q_a, Q_c) - Q_b) & \text{if } Q_c(t) > Q_{spillback} \text{ and } \max(q_a, q_c) > q_b \\ \Psi_{a\to b}(t) = h(Q_a(t) - Q_b(t)) & \text{if } Q_a(t) - Q_a(t) > 0 \text{ and } Q_c(t) < Q_{spillback} \\ \Psi_{a\to b}(t) = 0 & \text{otherwise} \end{cases} \quad (8.5)$$

As long as the probability of spillback is small, the extra delay given by this phenomenon will be also small. If on the other hand there is a non-negligible chance that the green time is not sufficient to clear the queue at the accumulation lane, the adjacent lane will also reduce in some cases its capacity, creating extra demand on the other lanes, as it will be shown in the case study section.

8.3.4 Two lanes example

The estimation of queues in a two-lane road can be computed using the Markov model and using formulas (8.2)-(8.5). To show the dynamic evolution of lane changing behavior a constant demand for the whole evaluation period is assumed in this example.

Flows are assumed to drive with 70% of vehicles on the right lane. Saturation flow is set to 1800 vehicle per hour, while cycle and green time are set respectively to 60 and 24 seconds. The accepted gap for a vehicle is expressed in time, and set to 3 seconds.

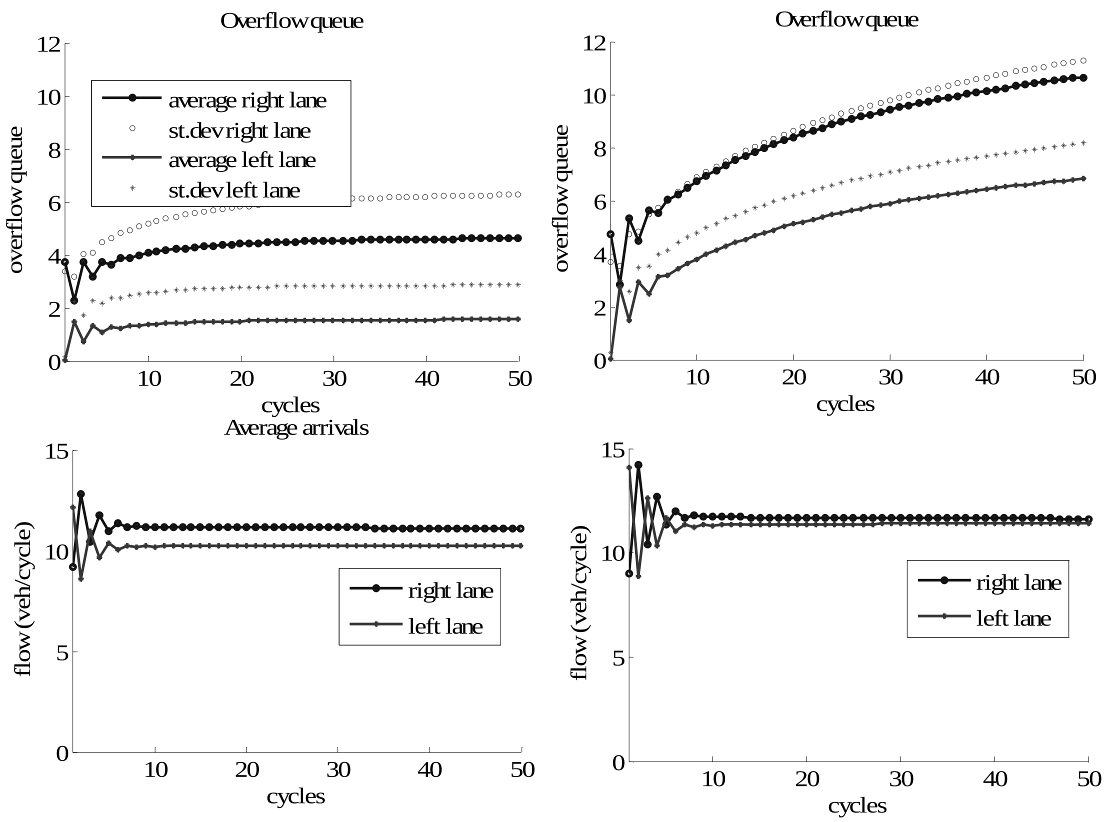

Figure 8.4: numerical examples of the expectation value of the overflow queue and flow distribution in time for x=0.96 and x=1

Figure 8.4 shows (top pictures) the evolution of the average and the standard deviation of the queues at the two lanes. The bottom pictures show the lane-changing behavior expressed by the evolution of the split rates in time. Left figures show the condition when the aggregated degree of saturation of the signal is set to 0.96, while the right pictures show the simulation with x=1. Because of lane changing possibility both lanes reduce their average overflow queue with respect to the single lane case. Standard deviation changes accordingly. The reduction of the variability of arrivals reduces both expected value and standard deviation of the queue.

Due to gap acceptance limitation, the two lanes are still not having the same demand. When the demand increases, here computed for a degree of around 1, lane changing also increases, since intention to change lane increases and flows tend to be equally distributed. When an equal distribution is met, queues tend asymptotically to increase with the same behavior.

Final remark should be made in the lane changing model. The proposed lane changing model results shown in the picture depend on the choice of the parameters for the lane-changing aspiration, the gap acceptance, and the dynamic queuing models. The calibration of these parameters is needed for its practical use. The parameters chosen for the example are only for illustration purposes and do not pretend to be realistic. For a more detailed description of the model calibration one can refer to (Hoogendoorn 1996) or (Ngoduy 2006).

8.3.5 Three lanes example and spillback effect

The presence of an exclusive turning lane, as displayed in figure 8.5, can influence the distribution of overflow queues among the lanes of a road section. In fact, the variability of the queue length can be such that spillback from the exclusive turning lane can reduce the capacity of the adjacent lane, forcing the vehicles driving on this lane to try and move to another lane in order to avoid an extra delay.

The following example refers to a total demand on the two lanes upstream the intersection of 2000 vehicles per hour. A right exclusive lane is considered with green time of 10 seconds and a maximum number of vehicles, above which spillback occurs, of 5 vehicles. Saturation flow is set to 1800 vehicle per hour, while cycle and green time are set respectively to 60 and 24 seconds for the straight through direction while only 10 seconds are given to the right turning direction. The distribution among lanes is set again to 70 and 30% respectively for the right and left lanes. Figure 8.5 shows the results when the percentage of vehicles going to the exclusive turning lane is set to 20% and 25% of the flow in the right lane.

As long as the chance to have spillback is small, the effect on the two adjacent lanes is small. The two left pictures are referred to a degree of saturation of 0.8 on the exclusive turning lane. If instead spillback is very likely to occur, as in the case of the example represented by the two pictures on the right side, the scenario changes consistently since a large part of vehicles will tend to move to the left lane.

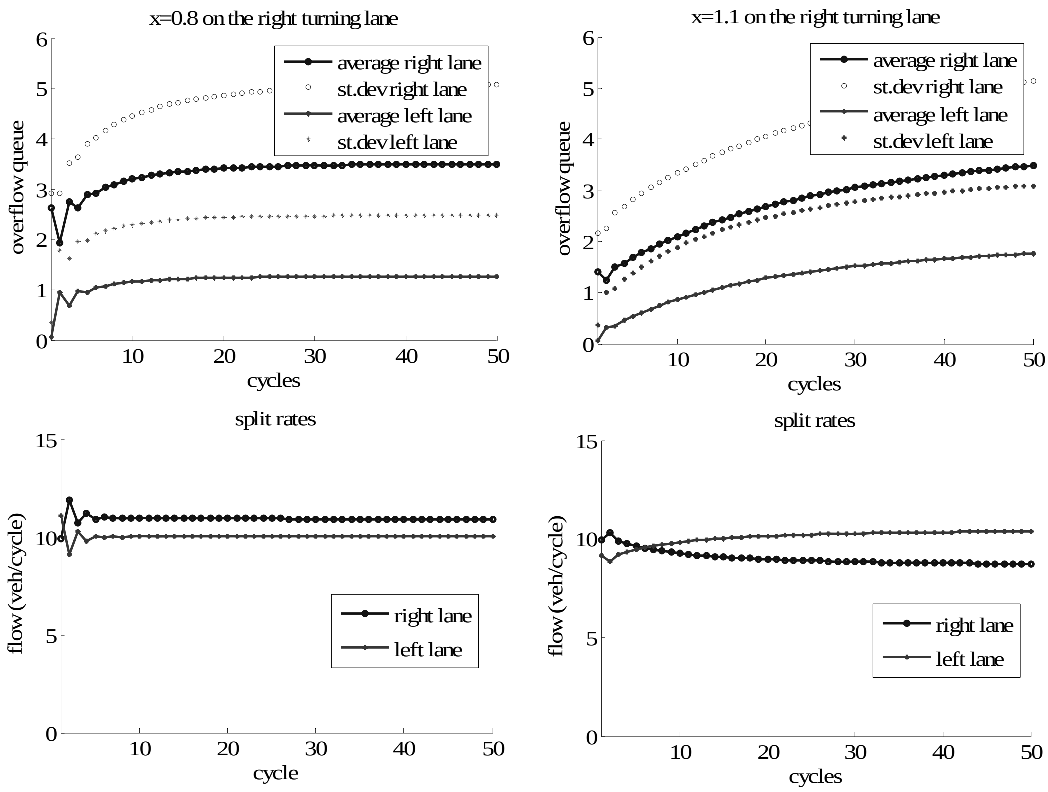

Figure 8.5: numerical examples of overflow queue and flow distribution in time for a percentage of right turning lane vehicles of 20% and 30%

The presented model can calculate also the probability that spillback occurs. For example for an average degree of saturation of 0.95 on the right turning lane, the chance of having spillback in a 15 minutes interval is nearly 35%, which is quite high.

8.3.6 Application of the multilane model in design problems

The proposed method allows one to estimates the queue length at each lane and to evaluate the queue length distribution especially taking into account the spillback effects. If the problem is to evaluate an existing infrastructure, the road manager can use this method to estimate the delay at each lane, since the geometry is already fixed. The method estimates the lane flow distributions according to the equal queue length principle. The difference from other models already proposed is the use of a dynamic queuing model and a gap-acceptance model. This method can be also used to optimize signal settings.

If on the other hand the road manager designs a new intersection or has the chance to modify the geometry, he can use this method to calculate the most convenient scenario. Since under the declared assumptions the lane changing behavior at the upstream section does not influence the queue evolution at the exclusive turning lanes, the design problem of flares can be restricted to the use of the single-lane queuing model.

Last remarks should be given to the underlying assumption of the method described in this section. A first remark regards the discussion whether a queue length-based lane changing approach is a valid approximation of real lane selection criterion of travelers at multiple service points. This represents certainly an approximation in some cases. For example drivers may not choose a shorter queue because of their awareness that one lane has a higher capacity or it favors some operations downstream (turning, merging or splitting lanes, lane drops). The equal queue criterion models drivers in these cases as “short sighted”. In other cases (a toll plaza for example) lanes (or service points) do not consistently make difference to drivers and the selection is more sensitive to large differences in queue lengths among lanes and this method may be the most appropriate. A solution to improve the model results can be combining the maximum queue length principle with e.g. minimum route travel time principle to combine these two effects.

Some microscopic simulation programs (e.g. AIMSUN, (Barcelo 2003)) use often a combination of these criteria. Microscopic programs frequently are used in the evaluation of queues and delays at multilane intersections, given their property of treating each vehicle as a physical entity and the possibility to simulate lane changing at the individual vehicle level. Appendix C shows microscopic simulations done using the program AIMSUN (Barcelo 2003) to estimate the changes in time of lane flows depending on the overflow queue length downstream. A strong relationship between lane flow distribution and queue length distribution was indeed found. The lane changing behavior of vehicles approaching an intersection was simulated under different demand conditions. The program confirms the queue-responsive lane changing behavior since lane flows tend to be equal. Overflow queues are also reduced since, looking at the arrivals near the intersection, the probability of a large number of arrivals is reduced.

Second remark regards the assumption of homogeneous traffic demand. This method does not consider the influence of heavy vehicles as reduction of throughput. This assumption should be stressed, since it is quite natural to expect that the presence of heavy vehicles affects the lane changing strategy of travelers. This investigation goes beyond the scope of this thesis and it is left for further research.

8.4 Time-dependent controls

The variety of dynamic control schemes designed and implemented in practice makes the development of a general Markov model very difficult. Nonetheless, the advantage to deal with each combination of queue lengths and number of arrivals in time is easily shown in the dynamic control context. In this section, a Markov model is formulated for actuated control signals.

8.4.1 Vehicle actuated controls

Actuated control phase plans are in general determined by the headway distribution of the arrivals at the intersection. The basic mechanism is to assign a green time unit (*unit extension*) when a vehicle is detected by the detection point. Green time extension stops when distance between two vehicles is larger than a certain threshold. This green time is usually constrained to be within minimum and maximum values, which are mainly determined by the geometry of the intersection. Since the stochastic nature of the arrivals, different headways and different number of arrivals can be observed from cycle to cycle. Therefore, the assigned green times and the delay incurred are variable too.

The assigned green times are thus variable according to the variability of queues forming during the red and green phases and to the variability of vehicle headways. If one considers that the queue formed during the red phase depends on the number of vehicles arriving during that phase and to the length of the red phase itself, which depends on the green time extensions given to all conflicting streams, the mathematical formulation of the expected travel time experienced by the travelers is quite complex. This section shows how this mechanism can be modeled using a probabilistic approach.

8.4.1.1 Vehicle actuated mechanism

Figure 8.6 explains how vehicle actuated signals work. The figure shows two loop detectors placed underneath the road pavement and at some distance from the stop line.

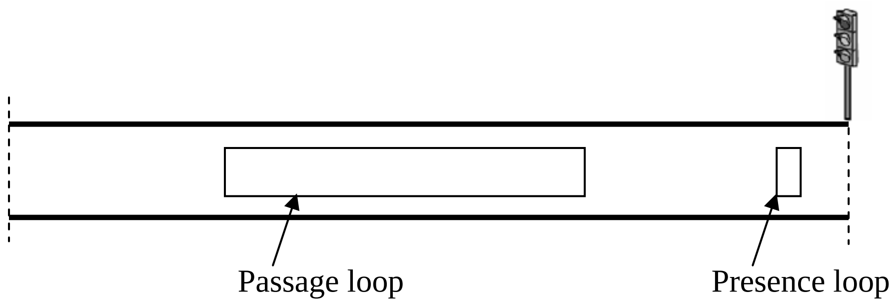

Figure 8.6: Loop detectors before a vehicle actuated signal

The closest loop to the traffic signal (*presence loop*) detects whether at least one vehicle arrives during one cycle length. If no vehicles arrive the signal will never turn green. The second loop (*passage loop*) defines how long the green phase should be. The mechanism is explained more in detail with Figure 8.7.

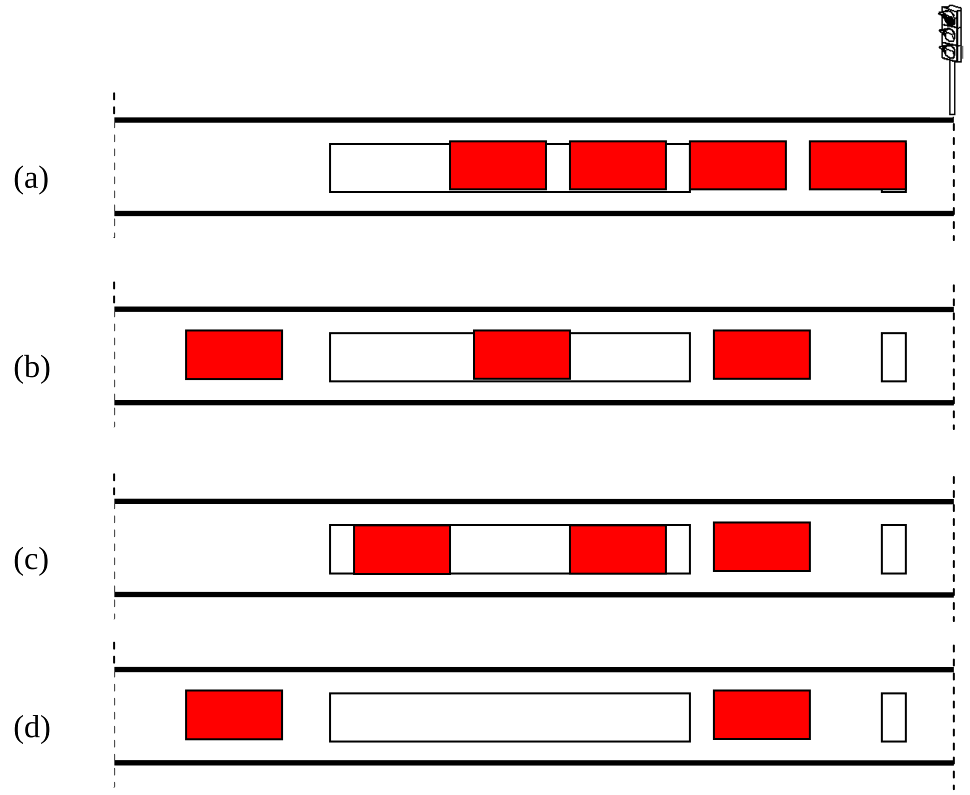

Figure 8.7: Vehicle actuated mechanism

Figure 8.7 (a) represents the case of a queue located over the passage detector. This can be observed during the red phase, when the queue builds up, and also during the green phase, when vehicles are being served but other vehicles reach the back of the queue in the meantime. As long as the queue is detected by the passage loop the green time is extended, unless this time is in between a minimum and a maximum value. Figure 8.7 (b) and (c) are two cases in which green time is still extended because vehicles arrive in short distance among each other. Figure (b) represents the case of only one vehicle detected within the unit extension, while Figure (c) shows two vehicles lying on the detector within the same unit extension. In both cases the green time will be extended of only one unit extension. Finally Figure 8.7 (d) shows when the distance between two consecutive vehicles is larger than the length of the detector. A timer counts the maximum time allowed to extend the green time further and if no vehicle arrives the signal will turn to amber. Other vehicle actuated mechanisms consider a fixed time extension for each vehicle counted, independently on their time headway. This mechanism is therefore simpler to be formulated since green time extension becomes only function of the number of arrivals within a cycle.

8.4.1.2 Computation of green time to clear the queue

Let $\overline{\tau}$ be the unit extension, assumed known and constant. This unit extension is determined by the length of the passage loop, the speed at which vehicles drive on it and the time extension of the timer. The length of the loop detector is usually fixed in such a way that the queue formed during the red phase is completely served, avoiding that the signal changes while the queue is not yet fully served or unless the maximum green extension is met. If $Q^r(\tau)$ is the total number of vehicles queuing up and waiting at the signal at the cycle τ during the red phase $r(\tau)$ the expected green time given to serve these vehicles is given by:

$$g^r(\tau) = \begin{cases} g_{\min} & \text{if } Q^r(\tau)/s \le g_{\min} \\ Q^r(\tau)/s & \text{if } g_{\min} \le Q^r(\tau)/s \le g_{\max} \\ g_{\max} & \text{if } Q^r(\tau)/s \le g_{\max} \end{cases} \tag{8.6}$$

where s is the saturation flow, assumed here known and constant. This value does not consider yet eventual vehicles arriving during the green phase and while the formed queue is served. The queue length $Q^r(\tau)$ is determined by the number of vehicles arriving only during the red phase. If one assumes uniform time headway, the expected green time is computed with Formula (8.6) using simply the average flow rate.

The probability distribution of green times at one traffic stream i is computed knowing the probability distribution of arrivals at stream i and the distribution of red times at the previous phase, which depends on the distribution of arrivals at all conflicting streams[2].

If $a_i(\tau)$ is the arrival rate (in vehicles per second), and $r_i(\tau)$ is the red time at the previous cycle one can compute the probability of a certain number of vehicles k queuing up during the red phase of length ρ as:

$$\begin{aligned} \Pr(Q_i^r(\tau) = k) &= \int_{\rho=r_{\min}}^{r_{\max}} \Pr(Q_i^r(\tau) = k \mid r_i(\tau) = \rho)\, d\rho = \\ &= \int_{\rho=r_{\min}}^{r_{\max}} \left(\Pr(a_i(\tau)\cdot\rho = k)\cdot \Pr(r_i(\tau) = \rho) \right) d\rho \end{aligned} \tag{8.7}$$

The formula is obtained by assuming that the arrival rate and the red time are independent stochastic variables. The integrals are assumed computed within a maximum

2 In practice the computation determines first the critical paths, which are all the conflicting streams with the largest demand.

r_{max} and a minimum red time r_{min}. The formula to compute the probability of red time length is given later in this section.

The probability of a green time $g^r(t)$ needed to clear the queue at the end of the red phase to be a value l is given by the following condition:

$$\Pr(g_i^r(\tau) = l) = \sum_{k=[s \cdot l]} \Pr(Q^r(\tau) = k) \qquad (8.8)$$

While clearing the queue formed during the red phase, other vehicles may join the queue. Formula (8.9) computes the probability for these vehicles similarly to Formula (8.7). The probability of red time is simply replaced by the probability of green time computed with Formula (8.8):

$$\Pr(Q_i^g(\tau) = k) = \Pr(Q_i^g(\tau) = k \mid g_i^r(\tau)) = \int_{\varsigma = g_{min}}^{g_{max}} \left(\Pr(a_i(\tau) = k \mid \varsigma) \cdot \Pr(g_i^r(\tau) = \varsigma) \right) d\varsigma \qquad (8.9)$$

Accordingly, Formula (8.8) is adapted to compute this extra queue with the following formula (8.10):

$$\Pr(g_i^g(\tau) = l) = \sum_{k=[s \cdot l]} \Pr(Q_i^g(\tau) = k) \qquad (8.10)$$

The probability of green time due to all vehicles in queue g^Q is thus given by the following relationship:

$$\Pr_i(g_i^Q(\tau) = m) = \sum_{k+l=m} \Pr_i(g_i^r(\tau) = k) \cdot \Pr_i(g_i^g(\tau) = l)) \qquad (8.11)$$

The total queue to dissipate within this time is computed accordingly:

$$\Pr_i(Q_i(\tau) = m) = \sum_{k+l=m} \Pr_i(Q_i^r(t) = k) \cdot \Pr_i(Q_i^g(t) = l)) \qquad (8.12)$$

8.4.1.3 Green time extension due to short arrival headways

Apart from the green time assigned to clear the queue, one should take into account that the green phase is extended as long as a vehicle passes the detector within the unit extension, which it can happen also when the queue has been fully served but vehicles are still arriving in short distance. To account for this extra-time one can use the distribution of arrival headways instead of the number of vehicles arriving within the time period.

The Poisson distribution can describe for example the probability of observing n arrivals in a period from 0 to t with the following expression:

$$\Pr_n(t) = \frac{(\lambda \cdot t)^n}{n!} \cdot e^{-\lambda \cdot t} \quad (8.13)$$

This equation gives information about how the probability is distributed over a time interval in terms of number of vehicles. In a sequence of n arrivals one can observe vehicles passing with a random headway distance. If no arrivals are observed within a time $\tau < \bar{\tau}$ the signal will switch to amber. This probability is given by the following Formula (8.14):

$$\Pr_0(t) = e^{-\lambda \cdot t} \quad (8.14)$$

This equation shows that probability that no arrival takes place during an interval from 0 to t is negative exponentially distributed. Given the unit extension, one can compute the probability that no vehicle will be detected (around 0.8 in the example shown from the graph for a 4 seconds unit extension).

If one computes the probability distribution of a sequence of n vehicles at times $0 < t_1 < t_2 < ... < t_n = t$ the probability of observing this sequence with $t_2 - t_1 < \bar{\tau}$, $t_3 - t_2 < \bar{\tau}$, etc. is given by the following Formula (8.15):

$$\begin{aligned} \Pr(t_{extn} = t) = \sum_{n=0}^{n_{\max}} & \Pr(t_1 < t_2 < ...t_n = t) \cdot \Pr(n,t) \\ s.t. \quad & t_1 < \bar{\tau}, \\ & t_2 - t_1 < \bar{\tau}, \\ & ... \\ & t_n - t_{n-1} < \bar{\tau} \end{aligned} \quad (8.15)$$

Even if green time extension is needed, one should compute the probability that this extension time is actually available. The probability of a certain number of seconds available for eventual green time extension before the maximum green extension is easily derived from the probability of green time due to the formed queue, Formula (8.12). The probability of having an extension of exactly t seconds (expressed as an integer value) is given by the following relationship:

$$P_i(g_i^e(\tau)=t)=\sum P(t_{extn}=t)\cdot P(g_{\max}-g_i^Q(\tau)\geq t) \tag{8.16}$$

The average total green time is finally given by computing the relationship $g^{tot}(\tau)=g^Q(\tau)+g^e(\tau)$.

The probability of a total green time $g^{tot}(\tau)$ is thus given by the following relationship, analogous to Formula (8.11):

$$P_i(g_i^{tot}(\tau)=m)=\sum_{k+l=m} P_i(g_i^Q(\tau)=k)\cdot P_i(g_i^e(\tau)=l)) \tag{8.17}$$

Green times are computed using this method for each flow stream of the intersection and the total cycle is computed by summing up all green times at each conflicting stream, together with the corresponding lost times. Knowing then the green times and the cycle length one can finally compute the uniform delay component using for example the delay model as in the fixed time case (Chapter 4). If the green time assigned during the previous cycle is smaller than the maximum green extension no overflow queue is supposed to be present and only the arrivals during the red phase should be served.

8.4.1.4 Computation of the overflow queue length

Overflow queues are likely to occur only when the intersection is oversaturated and the maximum green extension is met. If the signal assigns the maximum green extension, one should calculate the eventual overflow queue, which will have to wait for the next green phase. Overflow queue occurs then only if $g^Q(\tau)=g_{\max}$. In this case the overflow queue is computed by formula:

$$Q_O^i(\tau)=Q_i(\tau)-g_{\max}\cdot s^i \tag{8.18}$$

with s^i the assumed saturation flow of the road section i. The corresponding probability is computed by the following formula:

$$P(Q_O^i(\tau)=q)=\sum_{[k-g_{\max}\cdot s^i]=q} P(Q_i(\tau)=k) \tag{8.19}$$

8.4.1.5 The effect of overflow queues on green times

Since an eventual overflow queue should be cleared in the next green phase, Formula (8.8) should also consider that, apart from the arrivals, also the eventual overflow queue should be served. Formula (8.8) is thus reformulated as:

$$P(g_i^Q(\tau)=l)=\sum_{(k+q)/s=l} P(Q_i(\tau)=k)\cdot P(Q_O^i(\tau)=q) \quad (8.20)$$

8.4.1.6 Computation of red time probability

Last step to compute Formula (8.8) is to derive the probability distribution of red times at the previous cycle. The red time is determined by the sum of all the green times given to the conflicting streams and the total time lost of the signal:

$$r(\tau)=\sum_{j\neq i} g_j^{tot}(\tau-1)+TL \quad (8.21)$$

The corresponding probability of a red time to be a certain value $r(\tau)=s$ is thus computed with the following formula:

$$P(r(\tau)=s)=P(\sum_{j\neq i} g_j^{tot}(\tau-1)+TL=s) \quad (8.22)$$

Assumption should be made on the initial red phase probability and on the initial overflow queue in order to compute the distribution in time. Supposing an empty signal, initially one can fox as initial value the minimum red phase.

8.4.1.7 Numerical example

In the following of the section the vehicle actuated control is modeled with the probabilistic approach in a simple case of two traffic streams crossing an intersection as the simple scheme in Figure 8.8 displays. Let a_{AB} and a_{CD} be the average flows (expressed in vehicles per second) and $s_{AB}=s_{CD}=1800[veh/hr]$. Let assign the first green time to the direction AB and an initial red time $r^{AB}(0)=30s$. Initial overflow queue is zero $Q_O^{AB}(0)=0$. The total lost time of the intersection is assumed $TL=12s$ and the unit extension is $\bar{t}=3s$. Finally minimum and maximum green times are respectively set to $g_{\min}=10s$ and $g_{\max}=60s$. Let assume the arrival rate distributed as Poisson and let the process have a stationary arrival rate for a period of 30 minutes.

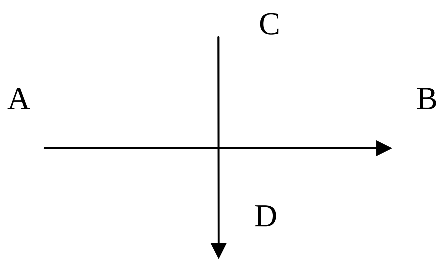

Figure 8.8: Two conflicting streams example

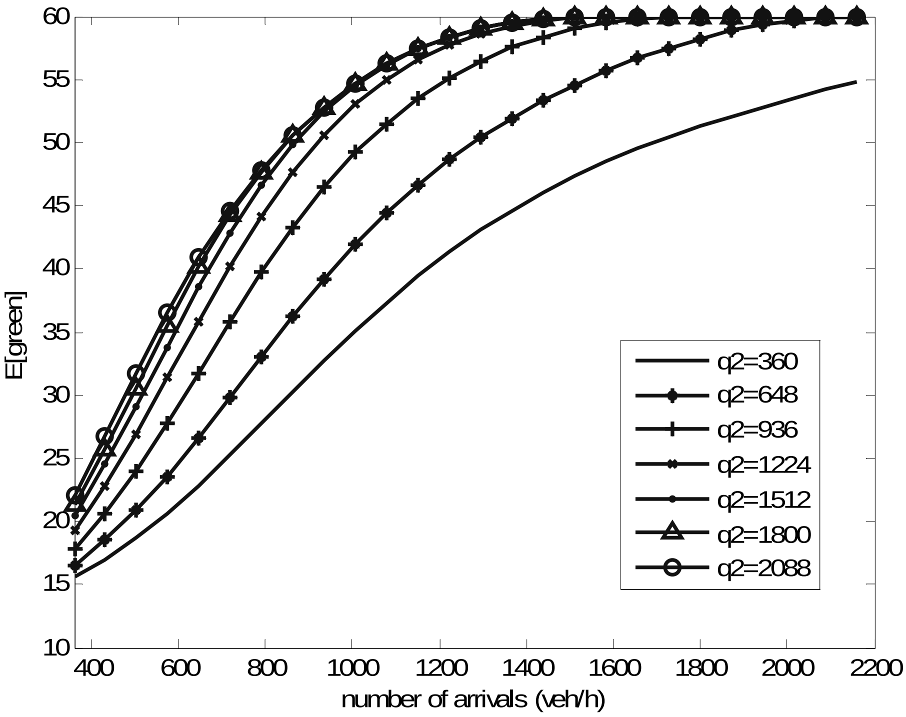

Figure 8.9: Expected green time length for stream AB for different demand conditions

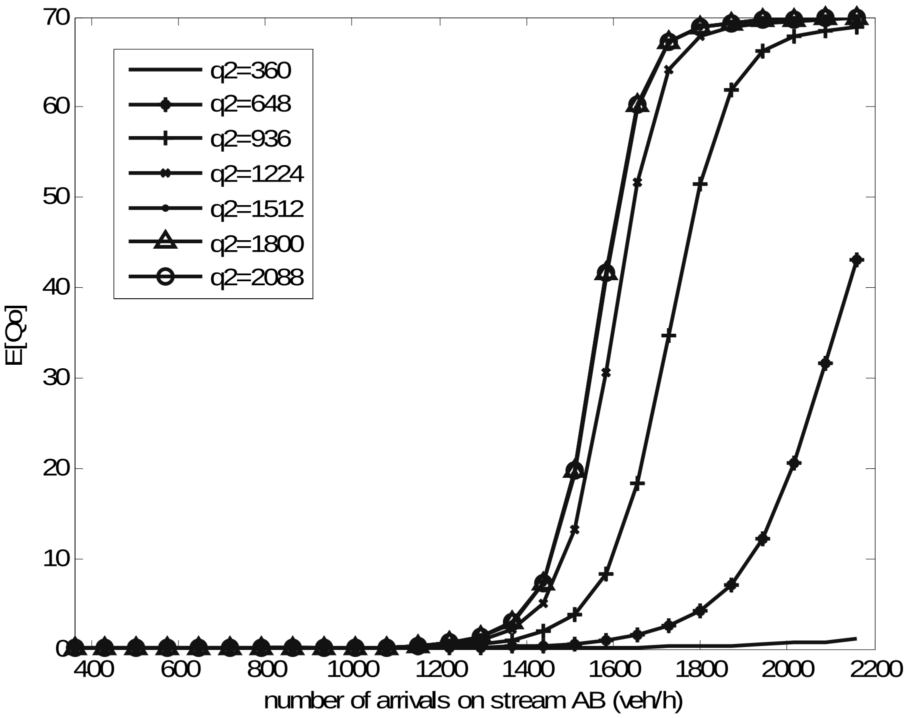

Figure 8.10: Expected overflow queue for stream AB for different demand conditions

Using Formulas (8.6)-(8.22) one can compute at each time step the probability distribution of green times and the eventual overflow queue in time. Expected green times and overflow queues have been computed for each couple (a_{AB}, a_{CD}), each ranging from a value from 0.1 to 0.6 of the saturation flow. Figure 8.10 shows the expectation value as function of the couple (a_{AB}, a_{CD}) at the end of the period of stationary conditions.

The expectation value is the minimum (or guaranteed) green time only when both flows are zero. The green time is sensitive to both the increase of each flow stream, but it has a steeper increase if increasing a_{AB} of one unit. The expected value is equal to the maximum green time for a large overall demand. Overflow queues are more likely to occur in these conditions, as figure 8.11 shows, while its value is nearly zero for low values of the demand, especially if a_{AB} is small. When one of the two streams has a very large demand overflow queue is likely to reach the maximum value assumed in the example (70 vehicles) within the total period of analysis.

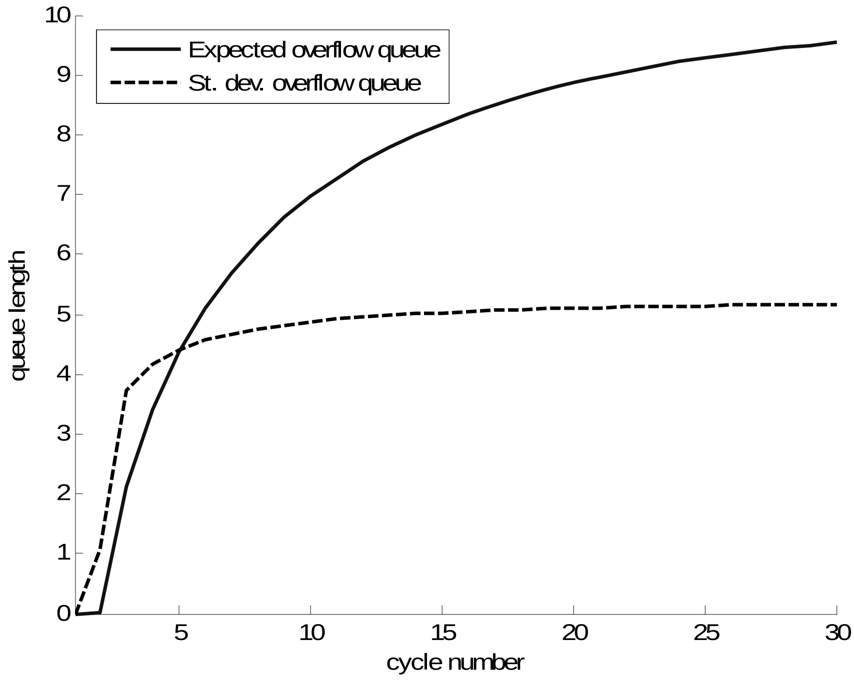

Figure 8.11: Expected overflow queue and standard deviation in time

The overflow queue in conditions of moderate saturation starts assuming a strong dynamic character. Figure 8.11 shows the example of $a_{AB} = 0.5 \cdot s_{AB}$ and $a_{CD} = 0.2 \cdot s_{CD}$.

8.4.2 *Comparison of pre-timed and vehicle actuated controls*

The formulation of the vehicle actuated control mechanism in a probabilistic fashion enables one to compare the performance of such controls with pre-phased controls based on the Markov model for fixed-time controls. Table 8.1 compares the results of the peak hour example computed previously in this thesis in terms of expected cycle length, expected delay and its standard deviation. The optimal cycle lengths set with the pre-timed method are similar to the average cycle length computed for the vehicle actuated control method only for very low demand conditions. On the other hand, pre-phased cycle lengths are considerably longer than the average cycle length of the vehicle actuated control scheme when the demand increases and delays are around 50% larger. Standard deviation is relatively small for both methods. Due to its flexible mechanism and the opportunity to adapt the signal settings on the arrival distribution of vehicles at the signal, vehicle actuated controls have been shown to be a more efficient control method with respect to the optimization of cycle times.

Table 8.1 – Assigned demand and resultant average delay with Markov chain and Webster methods

Flow (veh/h)	**600**	**700**	**750**	**800**	**750**	**700**	**600**	**500**
Vehicle actuated control								
Average cycle (s)	57	65	75	85	75	65	57	51
Expected delay (s)	14.7	16.5	18.5	20.9	18.5	16.5	14.7	13.2
St. deviation (s)	5.4	5.9	6.4	6.9	6.4	5.9	5.4	5.0
Fixed time controls based on total delay minimization								
Optimal cycle (s)	58	90	126	126	126	86	58	50
Expected delay (s)	19	30.1	31.1	31.8	30.9	30.2	19.5	14.1
St. deviation (s)	4.5	4.5	4.4	4.3	4.4	4.5	4.5	4.8

8.5 Summary

This chapter stressed some hypotheses that characterize the Markov model for single-lane fixed time controlled intersections described in Chapter 4. Three directions have been followed: arterial corridors, multilane sections, and time-dependent controls.

The hypothesis of Poisson distribution, assumed for isolated intersections, does not well represent the arrival distribution when signals are at short distances one to another. The filtering effect of upstream signals influences the maximum number of arrivals observable within a cycle to the downstream intersections. To account for this effect in the Markov model one can simply assume a maximum arrival, which depends on the flow streams that converge to the downstream signal. Nevertheless, the hypothesis of isolated signal represents an upper bound for the computation of overflow queues and stochastic delays.

One-lane sections work differently from multilane sections, since in the latter vehicles have the possibility to change lane. To account for this effect the Markov model has been combined with a lane-changing model. By doing so, the distribution of arrivals has been shown to have a dynamic character, according to the dynamic character of the overflow queue length. Furthermore, the Markov model at multilane sections allows one to account for spillback effects, which is useful information for a correct estimation of delays and for the design of exclusive turning lanes. Microscopic simulation programs use also similar lane changing criteria, as it is described in Appendix C. An improvement in the model would be combining the queue-responsive criterion with a route travel time criterion, to account for eventual effects of other delays upstream,

Finally the assumption of fixed control settings has been relaxed by formulating a probabilistic model of vehicle actuated controls. This approach allows one to compute the probability of green time extension depending on the variability of arrivals and their headway distribution in time. The probability of overflow queues is computed accordingly. The knowledge of these two elements enables the computation of the expectation value of delays and their distribution in time as it was done for the fixed control case. The complexity of the formulation is still limiting its application; heuristic formulations can be searched as it has been done in Chapter 6. Moreover the validation of the model should be done in future research, for example using microscopic simulation as it was done in Chapter 7 for the fixed control case. Nevertheless, this formulation represents, to the author’s knowledge, the first delay estimation model, which accounts for both the effects of the dynamic and the stochastic character of flows to travel times and queues at vehicle actuated controls.

9

Recommendations and application perspectives

9.1 Introduction

The Markov model presented in Chapter 4 and the Van Zuylen-Viti models presented in Chapter 6 improve the delay estimation and prediction of queues and delays at signals with respect to the analytic models presented in Chapter 3 in two ways:

- These models capture the dynamic and stochastic character of delays in a more realistic fashion than the existing analytical models by computing overflow queues under weaker assumptions. Computing of the probability of an initial overflow queue at each sub-period allows the traffic analyst to have estimates of route costs in time.
- The relationship between overflow queues and delays described in section 4.6 provides an explicit estimate of overflow queues and their effects on travel time variability. This information is important for estimating e.g. the route travel time reliability, the level of service of an intersection and for having a measure to compute the travelers' utility in an uncertain scenario.

Chapter 8 has given also an idea of the modeling power of the probabilistic approach. This approach has in fact been applied to evaluate the effects of upstream signals, multiple service points and dynamic controls in the dynamic and stochastic behavior of queues. These represent only examples of the opportunity given by this approach.

This chapter briefly discusses the potential areas of applications of the newly developed models. There are three main areas of applications for these probabilistic models. Figure 9.1 shows these three areas in a scheme, together with the more specific application problems within these areas.

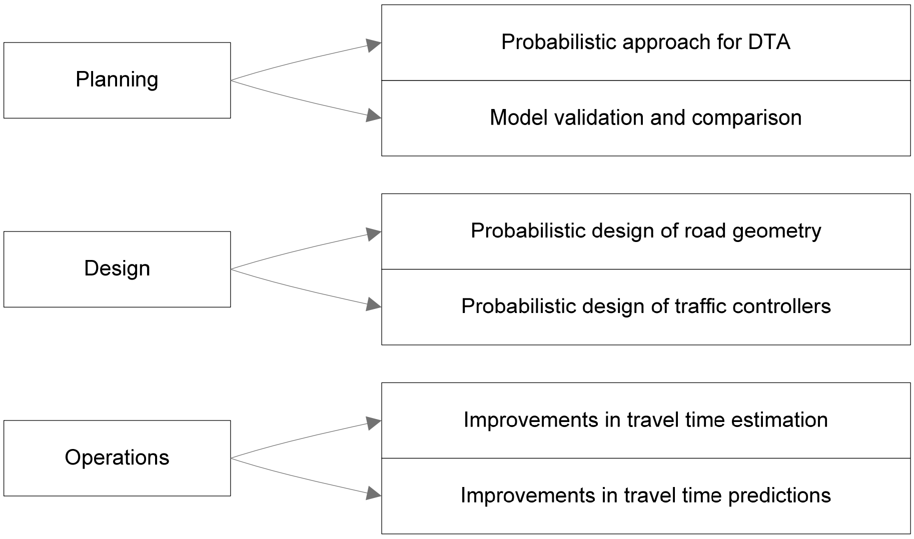

Figure 9.1: Fields of application for the probabilistic models

Long term travel time predictions, used for example in design and planning problems, need delay models that simulate the traffic in its most likely states in the current network and under different hypothetical conditions (e.g. future growth of the demand, changes in the network infrastructure or in the signal control plans etc.).

Despite the growing power of computers and therefore the growing interest in microscopic programs to simulate how the traffic propagates along a network, analytic travel time models are still considered a valid approach for design and planning purposes since the work of Davidson (Davidson 1966). For a general overview on this discussion one can refer to (Rose 1989) or (Akcelik 1991).

The newly developed models contribute to having better insight into this system in three ways; it can be applied to estimate route flows, to design or manage the supply system and to estimate the (expected) travel times. The following sections give some examples of these applications.

This chapter is structured as follows. The next section gives some possible applications of the probabilistic models in planning problems, while section 9.3 covers the area of the network design problems. Section 9.4 shows how the model can be applied in short term travel time predictions. Finally section 9.5 gives a synthesis of this chapter.

9.2 Applications in planning problems

The estimation of route flows along a network, both in terms of time and space, strongly depends on the assumed travel time function, together with the utility function and selection criterion assumed for the travelers. If the applied travel time function is not able to catch the dynamics of traffic, an incorrect distribution of flows may be obtained.

The newly developed models improve the estimation of traffic flows by giving the opportunity to catch the dynamics of travel times depending on the dynamic character of the demand system. Both the probabilistic approach and the Van Zuylen-Viti models give the opportunity to represent the overflow queuing process in time with results that are consistent with microscopic simulations. This can be used for example to estimate the effect of peak period congestion to urban travel times and to evaluate the effects on the choices of the drivers, e.g. route and departure time choice, under the assumption that expected travel time (and its standard deviation) are determinants of travel behavior.

Another main improvement that can be granted is the opportunity to include the uncertainty in the choice process of travelers. Chapter 2 has shown the strong value stated by a sample of drivers to the uncertainty of travel times. Dynamic Traffic Assignment (DTA) processes based on deterministic travel times consider often the drivers to choose their alternatives of travel depending on expectation values. Stochastic traffic assignment procedures consider, on the other hand, a component of uncertainty by including a noise in the drivers' perception of travel times. This error component is often considered in practice depending on the driver and not on the variability of travel times.

The importance given by the road travelers to the variability of travel times shown from the web survey presented in Chapter 2 suggests that the application of DTA to an urban network would improve if travel time variability is considered in the choice process of the drivers. The probabilistic approach or the Van Zuylen-Viti expression for the standard deviation can be used in such estimation to have an estimate of this variability and therefore give consistent estimates of the travelers' route choices with uncertain travel times. To give an example a TSL experiment where urban roads are presented to the respondents may improve the validity of its findings if travel times experienced during the experiment are drawn from a realistic probability distribution of travel times.

The models developed in this thesis can also help in the estimation of the risk averseness of the travelers. Some results form the TSL experiment described in chapter 2 and Appendix C show the travelers to have strong risk averseness by preferring more reliable routes, but also to have a non-linear risk behavior with respect to travel times. The models developed can be used to estimate this risk averseness.

9.3 Application in network design problems

Analytic travel time functions can be preferred to microscopic simulation in the design of an urban traffic network for several reasons; here are listed a few:

- The design of a signal plan or the modification of a road section requires the evaluation of the effects of this intervention; therefore several set-ups for the signals and hypotheses on the kind of intervention to be done on the road geometry need to be analyzed;
- Dynamic demand-responsive controls like anticipatory controls (see e.g. (Taale 2003)) require long iterative processes to evaluate the effects on the demand and the consequent adaptation of the control settings to these changes;
- The computation of supply characteristics like e.g. the intersection capacity requires models that enable the analyst to evaluate the sensitivity of these characteristics to the state and the control variables, e.g. the demand and the signal control settings;
- The design of the road geometry requires also models that allow one to evaluate different types of intervention on the road infrastructure or variations in the road geometry (e.g. the length of a flare or the number of lanes per flow stream).

In practice, the majority of transportation planning and design problems at urban controlled networks deal with delays by considering their expectation values. This approximation may lead to several drawbacks; for example, signal optimization may improve considerably the efficiency of one signal control when the actual flows are close to the average, but, on the other hand, it can be likely that this optimization creates large delays in a certain area when larger demands occur, for example creating gridlocks and blocks to other intersections upstream. This phenomenon may create extra delays, which cannot be computed if only expected values are considered. Chapter 4 has shown that the Markov Chain model can be used as an alternative to the Webster’s formula to calculate the optimal cycle time.

The two main contributions given by the models developed in this thesis are therefore very appealing in the design of dynamic control plans, like pre-phased controls, or to evaluate the reliability of these control plans, for example by evaluating the probability of cycle failures and thus overflow queues to occur. Moreover, the introduction of a probabilistic approach for the computation of travel times opens up the opportunity to have a probabilistic planning and design of the road networks. This approach has been for example supported in the past for defining the efficient use of the motorway capacity ((Brilon 2005)), or the optimal design of a flare ((Tian 2006)). It is therefore a very appealing methodology for measuring the reliability of the network design.

To give a few practical examples of these applications one can think of introducing, in the evaluation of the performance of a signal, a probabilistic definition of the level of service. The Highway Capacity Manual considers for degraded level of services (levels D, E and F) the possibility that large fluctuations in the demand produce unexpected large delays. The opportunity given by the new methodologies to calculate the chance of these large fluctuations in the delay experienced by the drivers enables one to have a quantitative estimate of this chance to occur.

A second useful application of the probabilistic approach is in design problems that have particular regard to spillback effects from short lanes of from one link to other links in the network. The Markov method in fact improves the optimal cycle method by also considering the physical length of the queue together with its probability to occur in time. In the optimization of area controls, the area manager may be interested in setting the signals in such a way that the probability of spillback is low in some sections.

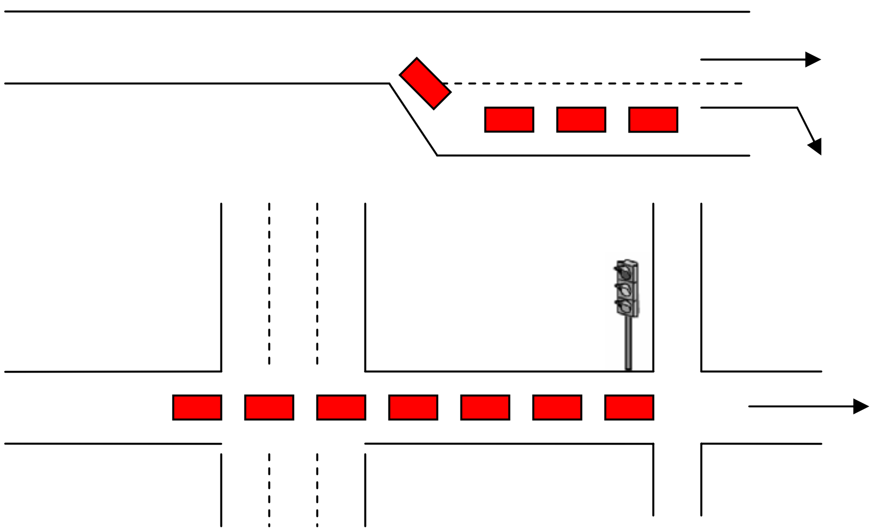

Figure 9.2: Examples of spillback effects

Figure 9.2 shows two cases where the traffic manager might consider spillback effects in the optimization of signals. The top figure may be the example of a short exclusive turning lane. The Highway Capacity Manual proposes to compute the optimal green and cycle lengths by considering the two lanes as separated and not influencing each other. In reality, if green time on the exclusive turning lane is not properly set, there can be a non-negligible chance that spillback occurs. The lower picture may represent a secondary road placed in short distance from a highly capacitated road. In both cases the road manager may apply the Markov method by simply computing the optimization of greens and cycles adding a constraint to the probability of a maximum queue length to occur.

Feasible solutions can thus be found by adding the condition that the probability of spillback does not exceed a pre-determined percentage of the total probability.

Normally the assumption in such problem is that the queue length has a Poisson distribution. Chapter 4 has shown that this assumption is incorrect and the shape of the probability distribution changes consistently from undersaturated to oversaturated conditions. The opportunity given by the Markov model to compute these probability distributions in time enables one to correctly estimate the probability of spillbacks and select the most opportune length for the links and the signal plan.

9.4 Application in travel time estimation and prediction

Accurate short term travel time predictions are fundamental component for many advanced traveler information and management systems. So far, the research on developing accurate methods to predict travel times has mainly focused on freeway analysis, while very little can be found on the urban context (Liu 2006). The main reason for a consistent difference between freeway and urban travel time predictions is the difficulty of including the effects of controlled signals upon the traffic flow.

Generally speaking, short term travel time predictions use measurements of the actual conditions of traffic (e.g. detectors, cameras) to give an estimate of how the system will behave a certain period of time afterwards. Since the collected data can give correct estimations of actual and past events, but no definite information is told about what can happen in the future, various methodologies have been proposed in the past to derive travel time forecasts from detector outputs (e.g. (Sisiopiku 1995), (Van Lint 2004)).

The newly developed models are useful also in this type of applications in many aspects. A few possible areas of application are here listed:

1. Model-based travel time prediction: short term travel time predictions in large networks require delay models that are able to predict how likely they will perform within a limited period of time. This information is useful e.g. in the implementation of ATIS systems and adaptive controls.

2. Data completion/validation: model-based models can improve data-driven models for two main reasons: 1) they can be used for the interpolation of missing data, which are typically scattered both in space and time and 2) they can be used in the data filtering process, for example detecting outliers.

3. Incident detection: in safety analyses it is often a difficult task to recognize when very long delays are due to traffic accidents and when these are simply due to the

stochastic nature of traffic. This information can be used e.g. in analyzing data or in real-time incident detection systems.

In general a travel time prediction is affected by three source of error: errors in the model (due to e.g. its simplifying assumptions), errors in the dataset (e.g. missing or corrupted data) and errors in the combination of the model with the data. Some recently adopted techniques (e.g. Kalman filtering, (Liu 2006)) try to minimize the last source of error, but its result strongly depends on the first two types of error. While reduction in the error from the data is obtained simply by using more accurate data collection methods (e.g. tracing each individual vehicle trajectory), reduction in modeling errors is obtained adopting models that use the least simplifying assumptions possible. The newly developed models are therefore improving the modeling power since they better catch the dynamic and the stochastic character of delays under very general assumptions for the demand and the supply systems. Moreover, the new models enable one to measure the modeling error by computing the uncertainty that affects its prediction. This information is vital in travel time estimation methods that hybridize model-based and data driven approaches.

In the first type of application the data collected from the current traffic status (or past traffic states) can be used to “refine” the prediction power of the models. Comparing the expected results computed with the probabilistic approach with data collected from any monitoring system, one can refine the model by e.g. having outcomes closer to the field data, or by having a better estimate of the relationship between the model parameters (flows, capacity, etc.) with the expected outcomes. Inversely, in the second type of applications the new models can be used to refine datasets, filling up missing data in space and time, or detecting outliers and “strange” data points.

9.5 Synthesis: the usefulness of a probabilistic approach

The newly developed models presented in this thesis, i.e. the probabilistic modeling approach presented in Chapters 4 and 8, and the Van Zuylen-Viti formulas developed in Chapter 6, give very important contributions to several areas of application in transportation problems. These areas of application are identified by three main classes: planning and design problems, traffic flow estimation problems and short term travel time prediction problems.

The most important improvement is certainly in their better estimation power with respect to the available analytic formulas for planning purposes (e.g. the HCM 2000 delay formula). It is recommended in these applications to use the new models since they give a better prediction of the dynamics of queues according to the dynamic and stochastic character of the arrivals and of the departures. Both the design of the signal plan and of the road geometry will improve using a probabilistic approach since it enables a probabilistic design of such characteristics and it enables the computation of the effects of spillback. Moreover, the probabilistic approach is a flexible methodology, which allows one to model accurately different types of control (e.g. pre-phased and vehicle actuated controls) and to select for each network scenario the most suitable type of control to implement.

The models presented in this thesis can also be recommended for applications in flow estimation problems like Dynamic Traffic Assignment problems, since these models improve their validity if one applies a delay estimation method, which enables one to catch the dynamics of travel times and therefore of the travel costs experienced by the drivers. Moreover, the knowledge of the variability of such queues (and consequently of the overflow delays) can be very important if the utility of a route alternative depends on the uncertainty of the travel time drivers may experience in their next trip.

Finally contributions can be found in the short term travel time predictions, since these predictions strongly depend on how general is the model in terms of correctness of its outcomes and of strictness of its simplifying assumptions. In this sense the newly developed models can help at refining the errors coming from the data collection method and to validate or complete such datasets. Finally, thanks to the knowledge of the variability of queues and delays depending on the variability of the arrivals and the departures, they can give a measure of the modeling error, which is fundamental information in hybrid model-based and data-drive travel time prediction models.

10

Conclusions

This chapter summarizes the research developed and presented throughout this thesis. The main contributions to the current state-of-the-art and practice in queuing modeling at isolated single-lane signalized intersections are briefly described in Section 10.1. The flexibility of this methodology is later shown in Section 10.2, in which the probabilistic approach is applied to more complex traffic flow processes (paired intersections, multiple lanes and dynamic control systems). The importance of developing the queuing models presented in this thesis is testified by the numerous applications that are described in Section 10.3. The importance of computing delays in a probabilistic fashion and the flexibility of the methodology adopted motivates further research in this direction, as indicated in Section 10.4. Finally Section 10.5 concludes this thesis with some recommendations.

10.1 Summary of research

In many cities, traffic congestion is observed systematically for a large part of the day, producing enormous economic and environmental losses as much as stress and dissatisfaction. A study in U.K. in 2000 reported that the total delay experienced in urban networks is 4 times larger than the total delay experienced on motorways during congested periods, and even 7 times larger in the London Area (Department of Transport 2000). There is general agreement that the largest part of the delay, due to congestion at urban networks, is caused by signalized intersections; this is justified both because non-signalized intersections are usually designed to regulate low-demand areas and because some signal service systems are often incapable of dealing with the variability and the dynamics of the arrival process (e.g. fixed timed controllers). The advances in computer technology and electronics have indicated a new direction of research towards a more efficient use of the transport network without the need of intervening on the physical infrastructure. Dynamic Traffic Management (DTM) measures are, among the so called Intelligent Transportation Systems (ITS), strategies that particularly have the scope of regulating the demand and the supply systems by either redirecting part of the demand to more capacitated alternatives of travel or by adapting the supply system to the actual demand. DTM strategies at signalized intersections are often designed to control the traffic flow by dynamically adapt the signal settings to the actual (or the expected) demand. To assess the impact of these strategies, and define optimal regulations, the traffic control planner needs models that correctly estimate the effects of these strategies on the network performances. To do so, deeper insight into the way the demand and the supply systems interact among each other is needed. Within these interactions travel times (and therefore signalized intersection delays) play a central role since they are determinants of the level of service of the supply and they influence the travelers' choices.

This thesis has the objective of improving the modeling and the understanding of the queuing and delay processes at signal controls; more specifically, it gives contribution to a clearer understanding of the dynamic and stochastic character of overflow queues at signal controls and the effects of these queues on the individual delay experienced by the drivers. This subject has been studied by many researchers in the past giving a vast amount of literature. This thesis presents a special approach to the delay problem by developing a new model for average or expected delay and its standard deviation. In particular, the effect of the variability of the arrivals on the formation and dissipation of overflow queues is thoroughly analyzed, giving insight into the queuing process especially in conditions of flows near capacity. Special attention is given to the variability of queues and delays, since travel time variability is fundamental measure e.g. to assess the travel time reliability, to evaluate the accuracy of information systems and to evaluate the travelers' satisfaction towards one alternative.

The objectives of this research are briefly summarized here:

- To give insight into the effects of the variability of arrivals to the dynamic and stochastic behavior of queues and delays at signals;
- To select a methodology suitable for simulating the effects of existing or newly developed management strategies in detail;
- To develop a new formula for the expectation value of the overflow queue length in time, which improves the delay estimation methods available in literature;

- To develop also a formula for the standard deviation of the queue length, which may improve planning and design problems that aim to estimate the reliability of a transportation network;
- To validate the newly developed analytical formulas.

The main result of this thesis is the development of new probabilistic models based on Markov Chain renewal processes (see Appendix A for a general overview), which explicitly account for the stochastic character of the elements involved in the traffic control system, i.e. the demand and the capacity. Starting from a methodology developed and applied already in past studies (e.g. (Van Zuylen 1985), (Olszewski 1990), (Wu 1990)) but limited to fixed controls and, apart from the work of Wu, to stationary demand conditions, new formulations for different types of controls (i.e. vehicle actuated controls), for multiple lanes, and for general traffic patterns were developed. Moreover, new analytic formulations for the expectation value and the standard deviation of overflow queues in time were derived from the results of the Markov model. The power of this methodology is testified by comparing its results to microscopic simulation, showing good consistency between them, and its flexibility is testified by the opportunity to apply this methodology to more complex scenarios.

10.1.1 Conclusions from empirical analysis and state-of-art review

Chapter 2 gives an empirical overview of the interrelation between the variability of the arrival process at urban routes and the variability of travel times. The causes and the effects of this variability are analyzed by looking at field measurements; differentiation is done between predictable variability and unpredictable (and therefore uncertain) one. Although estimation and prediction can be improved with enhancing the road monitoring system and by using detailed travel time models, only part of the demand fluctuations will be forecasted (e.g. day-to-day and within-day). The arrivals still remain stochastic, and this characteristic is especially affecting delays at signals that operate near capacity. This implies that, in terms of travel time uncertainty, signals should not operate too close to the signal capacity but spare capacity should be kept for the demand variations. This conclusion is in line with the concept of traffic efficiency used in the statistical concept of capacity at freeways used e.g. by Brilon ((Brilon 2005)) and the well know methodology to optimize traffic control settings of Webster ((Webster 1958)). Indirectly, this policy should be beneficial for travelers' information systems, since less variability implies a higher predictability, and, directly, for the travelers' satisfaction towards a route alternative, since drivers have been shown to be very sensitive to travel time uncertainty.

Travelers in fact take in consideration travel time variability sometimes even more than mean travel times. Especially when time constraints are involved (e.g. important meetings, delivery of goods within a time window), a road user is very concerned to the risk of long unexpected travel times and is willing to spend extra time on the road if this implies less risk of late arrivals. The management policy should therefore improve travel time predictability by reducing the sources of uncertainty. Moreover, travel choice models will improve their prediction properties if the costs for uncertainty are considered in the utility of a route and a departure time.

This thesis limits its area of attention to the delay incurred by vehicles at signalized intersections and the uncertainty around this measure; therefore, non-signalized intersections, roundabouts and uninterrupted facilities are not explicitly considered. Furthermore, this research is limited to motorized vehicles, in particular to passenger cars, since no regard is given to different vehicle classes and to the effect of one specific class to the others. Although the research presented in this thesis refers to this area, the models developed can be applied to all systems characterized by a stochastic arrival and service processes (so called G/G/n queuing processes, (Tijms 2003)). For example they can be applied to toll plazas, box offices, etc.

Since the seminal works of Beckmann et al. (Beckmann 1956) and Webster (Webster 1958) the delay at fixed-time controls has been assumed consisting of two components: the uniform and the stochastic delay term. The first expresses the delay as the arrival process is perfectly uniform, thus it is proportional to the chance of a driver to arrive at the signal during the green or the red phase, while the second computes the extra delay due to the random nature of the arrival process. These models are valid only for undersaturated intersections, since they assume the expected queue length within a cycle to be statistically in equilibrium. Models for the expectation values of the overflow queue length in these conditions were proposed to justify the existence of the stochastic delay component (e.g. (Miller 1963) or (Newell 1965)). Time-dependent queuing models for oversaturated conditions started to interest the research field from the work of Catling (Catling 1977), who proposed to use a simple linear relationship when the average degree of saturation is larger than 1. A discrepancy between the static and the deterministic queuing was then affecting the delay estimation. Approximate formulas were proposed to reduce the discrepancy between theoretical relationships and real-life using mainly heuristics (e.g. (Webster 1958), among various others for the static case, and (Akcelik 1980) for the dynamic case).

The heuristic foundation of the models presented in the literature study suggests that there is still no clear insight on the real dynamic behavior of these measures. The scientific forum agrees that a large source of modeling error is caused by inexact estimation of the overflow delay due to queues, and points to a lack in estimating the transitions between congested and uncongested conditions. This gap is primarily due to the large contribution given by the random nature of the arrivals within a cycle and among cycles. Clearer insight is therefore achieved if the relationship between the arrival distribution and the distribution of delays in time is well understood.

10.1.2 The old theory revised

Overflow queues are assumed to grow from cycle to cycle when the intersection is oversaturated, while in undersaturated conditions the overflow queue is assumed static and following e.g. the Webster's formula. The unrealistic character of this assumption becomes clear in the analysis of the transition between undersaturated intersections and oversaturated ones. Near saturation the overflow queue becomes infinite, according to formulas like Webster's, while it becomes finite and linear in time in oversaturated conditions. Heuristics have been used in the past to solve this discrepancy (e.g. (Kimber 1979), (Akcelik 1980)), proposing some time-dependent expressions for the queue and the delay processes based on a coordinate transformation, but no clear insight of how to derive an exact time-dependent expression has been provided so far.

Another assumption affecting the delay estimation is that both static and time-dependent queuing models are calculated starting from an empty signal, i.e. the initial queue is assumed zero. Only recently a third component has been considered necessary in the control delay estimation: the initial queue delay component, which expresses the extra delay caused by an initial queue that is larger than the random queue component. This component was adopted in the latest version of the Highway Capacity Manual ((TRB 2000)). The drawback still limiting this expression is in the way the dynamics of overflow queues are calculated. This model in fact assumes that the expectation value of the initial queue decreases according to the linear relationship proposed by Catling until its complete clearance. This linear model has been shown, when compared to the cycle-by-cycle Markov chain process, to underestimate the overflow queue. This error is particularly affecting the conditions of traffic near the capacity.

As a consequence of these limiting approximations, the analytic formulas developed so far are not applicable for example in peak period analyses (except for the model of Wu (Wu 1990)) since they do not properly model the transition between congested and uncongested condition, i.e. the expected behavior of decreasing queues. This affects for example the estimation of flows rescheduling their trip from peak to off-peak period and the estimation of queue clearance times.

The strong stochastic behavior of queues in conditions near capacity justified a stochastic modeling approach, the Markov Chain model, already adopted in the past by several authors (e.g. (Van Zuylen 1985), (Olszewski 1990), (Wu 1990)). Chapter 4 and Appendix A describe in detail this methodology, which has inspired consistently the models developed in this thesis. These approaches consider the queue length to be a dynamic process, i.e. a Markov Chain, where the probability distribution of overflow queues at one time step depends only on the distribution calculated at the previous time step, i.e. in the case of overflow queues, the queue state at one cycle depends only on the queue state at the previous cycle. However, the models proposed in the past are limited to evaluate the queue length probability distribution cycle by cycle while no insight is given on what happens within a cycle.

One main contribution of this thesis is the development of a Markov chain model also for the queuing process within a cycle, which extends the models developed in the past and helps understanding and explaining the overflow queuing phenomena observed in a cycle-by-cycle process. Chapter 4 has presented this Markov formulation; this model computes, in a probabilistic fashion, the expectation value of the queue length in time, growing during the red phase and being cleared during the green phase, which explicitly accounts for the effect of variable arrivals. While little difference is observable for relatively small volumes of traffic (i.e. below a degree of saturation of 0.8), the effects in time of the variability of the arrivals are rather considerable.

This model contributes to the state-of-the-art presented since it is an exact formulation, which considers explicitly the temporal effect of the variability of arrivals at any point in time, and not only from one cycle to the following as it was done previously. The new formulation of this random delay component is derived without any assumptions about the statistical properties of the arrival process, apart from the assumption that the arrival distribution is uniform over the whole cycle. An interesting research direction can be for example assuming a different arrival distribution at the signal, for example to evaluate the queuing process also when arrivals are influenced by some signal coordination.

The numerical value of the new random delay component can easily be computed and compared with the many approximate expressions developed in the past. More important than the possibility to calculate the random delay with a more general model than the existing ones, is the insight that the derivation gives in the process that causes the random delay. The assumption used by some authors that the queue should be represented by a step function due to the binary character of the arrival process, appears to be superfluous if the analysis is focused on the expectation value of the queue length. The stepwise character of the queue is transformed to a smooth function. The expectation value of the queue in the green phase shows a non linear character as soon as the tail of the probability distribution comes close to zero. This phenomenon causes the overflow delay also at undersaturated conditions.

The contribution of the new delay formulation is particularly visible at intersections operating near the capacity. Comparison with well known models of Webster ((Webster 1958)), McNeill (McNeill 1968) and Akcelik ((Akcelik 1980)) showed that these models are particularly inaccurate within this range. On the other hand, this method is limited in that it is solvable only numerically, i.e. no simple direct expression between degree of saturation and delay is derived. This makes it unsuitable for planning purposes in its present form. Two options are possible for its application: 1) as a validation for existing formulas of random delay or 2) as a valid data-generator to search for a simple formula that approximates this expression. This last approach has been followed in Chapter 5 for the evolution of Q_0 , the initial queue at the start of a red phase.

10.1.3 New insight into the dynamics of traffic at signals

The within-cycle delay model, combined with the cycle-by-cycle model presented in chapter 4 allows a full evaluation of the queue length and the delay distributions in time, regardless of what probability distribution for the arrivals is assumed. This enables one to estimate the delay incurred by the road users and the uncertainty around this estimation. This information, plugged in a network scenario, can improve the route travel time estimation and predictions and their uncertainty using model-based travel time prediction models or hybrid model-based and data-driven models (e.g. (Liu 2006)). Chapter 8 discussed the potential opportunity of applications in this direction. More importantly, this formulation can be applied in planning and design problems, since it is especially suited for long-term travel time predictions and to estimate the route flows along a network in a dynamic scenario. The knowledge of the variability of travel times can be useful information if a component in the utility of the drivers dependent on this variability is assumed in the estimation of route preferences. Moreover it is fundamental element in the estimation of the network travel time reliability.

The expected values of the queue length and its standard deviation have been found often to be of the same order of magnitude. The numerical evaluation using a Poisson distribution for the arrival process shows in fact the two elements to have nearly identical dynamics in time. If estimation is done assuming non-stationary flow conditions, but the arrival distribution is assumed stationary within sub-periods, the behavior of the queue length within each sub-period strongly depends on the queue length distribution at the end of the previous period and not only on its expectation value. To make an example of the implication of this remark one could think of two different applications for the model: if the model is used for planning purposes or for estimating flows through a dynamic traffic assignment process the delay estimated at a specific time (say, 8:30AM) will depend on the traffic pattern assumed in the past time periods. If on the other hand the model is used for short term predictions (e.g. (Liu 2006)) the information about the initial queue length is given by measuring the present state (e.g. by cameras); in this case the Markov model can be applied to predict the probability distribution on a future moment and it should be computed with a deterministic initial queue state.

This method has also quantified the error that affects standard analytical procedures like the HCM2000 ((TRB 2000)) to compute vehicle delays especially when the model should compute decreasing queues after an oversaturated period, which frequently occurs during peak periods. All available methods up to date underestimate vehicle delays, or flatten delays within an evaluation period, instead of computing more accurately delays for each cycle.

10.1.4 New approximate formulas for the expectation value and the standard deviation

The Markov process in Chapter 4 has inspired the development of a time dependent formulation for the expectation value of overflow queues, which well reproduces the results of the Markov model. This model is developed under milder assumptions than the heuristic formulas developed in the past (e.g. (Akcelik 1980), (Brilon 1990)) and it has been found to well reproduce also the results from microscopic simulations.

The ratio between standard deviation and variance influences the dynamic behavior of queues. This implies that an analytical expression for the standard deviation is also an important research issue. Therefore, an expression for the standard deviation has been also developed. To the author’s knowledge this is the first time-dependent expression for the standard deviation of the overflow queue length. The two analytic models have been called throughout this thesis the *Van-Zuylen-Viti* models.

These models have a broader area of use than the models presented in the literature study since they reproduce the evolution of the expectation value of the queues and their variability as function of time, without the necessity to fix an evaluation period but they provide estimates for every cycle. The models proposed can compute queues and delays assuming both uniform and non-uniform arrivals. This feature makes them suitable for dynamic and stochastic route choice processes, but also to make short-term prediction of expected waiting times. The standard deviation model is also very important to have an estimate of the uncertainty of a delay prediction, and it is valuable information for measuring the reliability of signalized route networks. Since the models have been developed for isolated intersections a multiplicative factor should be applied to the two models to consider the filtering effect of upstream signals when applied in a network.

10.1.5 Model Validation

Chapter 7 was dedicated to the validation of the Markov and the Van Zuylen-Viti models. Microscopic simulation has been used to validate the Markov approach and the analytic formulas proposed in this thesis. The alternative to use real life observations is practically impossible, since it is rather unlikely that one can observe in real life sufficiently long periods of stationary demand conditions. To operate the analysis several microscopic simulations were needed. Firstly the comparison was operated considering different scenarios, i.e. long period of stationary demand conditions to analyze the behavior towards equilibrium, behavior with different initial queues, behavior with non-stationary demand rates etcetera. For each scenario hundreds of microsimulations were necessary to obtain statistically significant estimates. The comparison was limited for situations that are close to saturation since the new models show consistent differences especially within this range. To the author's knowledge, this is the first study that attempts a comparison of models in a non-stationary demand rate scenario and that gives special importance to the behavior of overflow queue lengths from large initial values. Only a few studies have instead compared models in the way they deal with the variability of traffic at signals.

The results of the model comparisons are satisfactorily and the three models show the same dynamic behavior. The analytic function presented in Chapter 5 is therefore suitable for planning and design purposes and contributes to a better estimate and prediction of the signalized network performances. The results in terms of expectation value are nearly identical, while the standard deviation estimated with the Markov and the Van Zuylen-Viti models is slightly smaller than the microscopic simulation results in oversaturated conditions. The standard deviation shows especially larger values in the two models with respect to the simulated results of VISSIM (PTV 2003) especially when recovering from a large initial overflow queue at the start of a new sub-period. This has been explained by a more uniform behavior of the vehicles in the microscopic program with the increase of congestion due to the assumed car-following logic. It is not clear whether this different behavior is a weakness of the proposed models or of the microscopic car following behavior modeled in VISSIM. Future research will aim at solving this issue.

Simulation of the queuing process of an upstream signal has suggested that the arrival distribution at one signal inside an arterial corridor should take into account of the filtering property of signalized intersections. The variability of the arrivals is in these conditions limited in its largest values, reducing the chance of overflow queues. Simulation of the queuing process using the Markov model should therefore be done using this filtering property. The simulation has been done considering a single flow stream coming from the upstream signal, i.e. no arrivals from other directions (e.g. turning flows or secondary entrances) and it is therefore applicable in arterial corridors, where the flow rate of the main stream is considerably higher than the flows coming from other sources. Future direction of research on this issue should investigate the influence of flows coming from any upstream section converging to the signal and especially the effects of their variability.

The Markov and the Van Zuylen-Viti models are still valid for a single flow stream section and with fixed control. Although fixed time controllers are considered as old-fashioned, while most intersections are controlled by closed loop controllers, the traffic dependent control becomes nearly fixed time during peak hours. That makes a good model for delays of fixed time controlled intersections still important. Nevertheless, the model extensions presented in the next section show that the probabilistic modeling approach is a very powerful method to simulate the dynamics of a traffic flow process in more complex travel time estimation problems.

10.2 Model extension to general networks

Chapter 8 proposes the application of the Markov approach described in Chapter 4 in three directions: arterial corridors, multilane sections and time-dependent controls.

10.2.1 Effect of upstream signals

The hypothesis of isolated signal implies that the arrival distribution can be considered uniform in time, i.e. a vehicle is equally probable to arrive at every second of the cycle time. This hypothesis seems not suitable when simulating a sequence of signals. Signal coordination has been shown to affect only the uniform delay component, while it does not affect the overflow delay component. Upstream signals can affect the overflow delay component because of their filtering property. The filtering effect of upstream signals influences in fact the maximum number of arrivals observable within a cycle to the downstream intersections.

Microscopic simulations using VISSIM has shown that the arrival distribution profile in an arterial corridor (i.e. only the one flow stream is considered coming from the upstream signal) differs considerably from the isolated intersection case, and indeed the main difference is in the maximum number of arrivals per cycle. Widely applied overflow queuing models like the program TRANSYT ((FHWA 1984)) do not take into account this effect, since they consider all intersections as isolated.

To account for this effect in the Markov model one can simply assume a maximum arrival, which depends on the flow streams that converge to the downstream signal. Nevertheless, the hypothesis of isolated signal represents an upper bound for the computation of overflow queues and stochastic delays.

A numerical evaluation of an arterial corridor shows that both expectation value and standard deviation of overflow queues reduce considerably when an upper bound to the maximum number of arrivals is assumed. Below a certain value this characteristic becomes negligible, confirming the theoretical results of Newell ((Newell 1971)), who proposed to consider in an arterial corridor overflow queues to occur only at the first signal while no overflow queues should be considered in the following signals. This conclusion is although valid only below a certain number of arrivals assumed as maximum.

10.2.2 Effect of variable lane distribution

The hypothesis of single service point limits the applicability of the Markov model presented in chapter 4 in a multiple service point scenario. One-lane sections work differently from multilane sections, since in the latter vehicles have the possibility to change lane. To account for this effect the Markov model has been combined with a lane-changing model. By doing so, the distribution of arrivals has been shown to have a dynamic character, according to the dynamic character of the overflow queue length. Furthermore, the Markov model at multilane sections allows one to account for spillback effects, which is useful information for a correct estimation of delays and for the design of exclusive turning lanes.

The method estimates the lane flow distributions according to the equal queue length principle. The difference from other models already proposed is the use of a dynamic probabilistic queuing model and a gap-acceptance model.

If the problem is to evaluate an existing infrastructure, the road manager can use this method to estimate the delay at each lane, since the geometry is already fixed. This method is also suitable to evaluate or design the optimal length of a flare.

10.2.3 Markov model formulation for vehicle actuated controls

The hypothesis of fixed control settings is also limiting the applicability of the Markov model. It is common to use loop detectors to control the intersections according to the actual traffic conditions. Vehicle actuated control is frequently implemented in these areas but the estimation of the delay caused by this control type is not yet clearly understood.

The assumption of fixed control settings has been stressed by formulating a probabilistic model of vehicle actuated controls. This approach allows one to compute the probability of green time extension depending on the variability of arrivals and their headway distribution in time. The model assumes the green time subdivided into three main components: the green given to vehicles arriving during the red phase, the one assigned to the vehicles that arrive during the green phase but that have to stop because the queue has not yet been cleared, and the green time extension given to vehicles arriving in short headways while the queue has already been cleared.

The overflow queue is assumed to occur only when the maximum green time extension is met, and the probabilistic model computes its occurrence depending on the arriving flow rate of the stream and of the conflicting streams.

10.3 Future applications of the models

The central role of intersection delays at signalized intersections in many transportation problems justifies the many application areas where the models presented in this thesis can be placed. Here are listed only few examples:

1. Urban network travel time prediction: both probabilistic and analytic models may contribute to improve model-based travel time predictions in a dynamic environment. Moreover, the Markov and Van Zuylen-Viti models enable an estimate of the uncertainty of these predictions, which is important in measuring the inaccuracy of the prediction due to the variability of traffic flows. Information systems may be improved with this information.

2. Data completion/validation: an accurate model-based prediction, which enables one to deal with the dynamic and the stochastic character of traffic, can improve data-based travel time predictions for two main reasons: 1) it can be used for the interpolation of missing data, which are typically scattered both in space and time; 2) it can be used in the data filtering process, for example detecting outliers.

3. Development of approximate formulas: the probabilistic approach has been shown to be suitable tool to simulate a dynamic process accounting for the stochastic nature of the state and control variables. This approach has suggested new analytic formulas for the expectation value and the standard deviation of the overflow queues at fixed controls. A similar approach can be done for example for developing formulas with different assumptions (e.g. control type, arrival distribution, etc.).

4. Optimal design of signals: the models presented in this thesis improve the available models for design and planning purposes in two ways: 1) they catch well the dynamic behavior of the expectation value of the overflow queue and 2) they estimate the uncertainty of this value to occur. These two features improve the optimal design of signals, e.g. they improve the optimal pre-timed controllers, they improve the reliable design of network signals by computing the probability of spillbacks to occur etc.

5. Optimal design of road geometry: the knowledge of the dynamic and stochastic behavior of overflow queues at signals can be helpful in the design of the geometry of the signalized intersection, e.g. the optimal length of a flare, the optimal number of lanes per flow stream etc.

6. Application in flow estimation problems: the solution of an assignment process strongly depends on the assumed cost functions and on the assumed utility function of travelers. The application of travel time functions, which are not able to catch the dynamics of traffic and their propagation in time and space, may result in an incorrect distribution of flows in time and space. Given the strong travelers’ risk averseness towards uncertainty, flow estimation problems will improve if both the value of uncertainty for the travelers is estimated and if a travel time function that estimates this uncertainty is applied.

10.4 Recommendations

This last section aims at giving directions for further research following the study presented in this thesis.

10.4.1 Recommendations for model developments

In the development of the models presented in this thesis some assumptions were made. Further research can release these assumptions in the following directions:

- The assumption of Poisson distribution for the number of arrivals within a cycle and deterministic service rate is used as example throughout the whole thesis. Other distributions can also be used in the numerical evaluation of the overflow queue, probably modifying the shape of the expectation value and the standard deviation of the queue. This can lead to different approximate formulas than the Van Zuylen-Viti;
- The hypothesis of uniform arrivals within a cycle can also be relaxed in the estimation of delays at vehicle actuated controls. The platooning effect of upstream signals can strongly modify the results of the numerical evaluation of the within-cycle queuing process;
- The assumption of homogeneous traffic composition is also limiting the model accuracy. It is recommended to analyze the impact of variable traffic composition both to estimate the variability of the signal throughput and of the overflow queues;
- More complex delay estimations of dynamic control signals can be developed with the probabilistic approach, e.g. by including public transport priority, advanced network area controls etc.;
- Acceleration and deceleration effects at the start and the end of the green phase are not considered in the models. It is assumed that the waiting time stops at the moment that the queue in front of a car has left. To compute individual delays these effects should be considered.

10.4.2 Recommendations for model calibration/validation

The models described in this thesis have a strong theoretical background and therefore a few recommendations are needed for their application in practice:

- Validation has been done using microsimulation for its tractability. However, the soundness of the model approaches presented in this thesis will improve considerably if comparison with real-life data is done. Although field data is practically impossible to be obtained, some controlled experiments (i.e. recruiting several vehicles and drive them in a test site) could be done.

- The behavioral assumptions done to numerically evaluate the effects of overflow queues in the lane flow distribution have not been calibrated with field data. The sensitivity of drivers to lane change should be strongly dependent on the road intersection in analysis; also in this case a controlled experiment could be done.

- The vehicle actuated control model should be validated by microsimulation as it was done for the fixed control case;

10.4.3 Recommendations for model applications

The recommendations for the application of the models presented in this thesis are here listed:

- The models presented in this thesis are very appealing and should be used for design and planning purposes;

- It is recommended to select the arrival and the service rate distributions according to the values analyzed with historic data;

- These models can be used for short term travel time prediction if combined with data-driven models, as it was discussed in chapter 9;

- The models developed can be used in flow estimation problems, especially in Dynamic Stochastic Traffic Assignment processes, since they give one the opportunity to catch both the time-dependency and the uncertainty of queues and delays;

- The models presented are valuable tools for the estimation of the network and the travel time reliability.

The author envisages that the proposed probabilistic modeling approach opens new perspectives in the development of dynamic and stochastic travel time estimation problems and it is a suitable methodology for design, planning and reliability analyses.

Bibliography

Akcelik, R. (1980). Lane utilization and saturation flows. *Traffic Engineering and Control* **21**(10): 482-484.

Akcelik, R. (1980). *Time-dependent expressions for delay, stop rate and queue length at traffic signals*. Technical Report. Australian Road Research Board.

Akcelik, R. (1989). On the estimation of lane flows for intersection analysis. *Australian Road Research Bulletin* **19**(1): 51-57.

Akcelik, R. (1991). Travel time function for transport planning purposes: Davidson's function, its time-dependent form and an alternative travel time function. *Australian Road Research Bulletin* **21**(3): 49-59.

Akcelik, R., Rouphail, N. (1993). Estimation of delays at traffic signals for variable demand conditions. *Transportation Research Part B* **27**(2): 109-131.

Akcelik, R., Chung, E. (1994). Calibration of the bunched exponential distribution of arrival headways. *Road and Transport Research* **3**(1): 42-59.

Akcelik, R., Chung, E., Besley, M. (1997). Recent research on actuated signal timing and performance evaluation and its application in SIDRA 5. *67th Annual Meeting of the Institution of Transportation Engineers*, Boston.

Akcelik, R., Besley, M. (2001). Microsimulation and analytical methods for modeling urban traffic. *Conference on Advance Modeling Techniques and Quality of Service in Highway Capacity Analysis*. Truckee, California, USA.

Akcelik, R. (2002). *aaSIDRA Users' Manual*.

Akcelik, R. R., N. (1993). Estimation of delays at traffic signals for variable demand conditions. *Transportation Research Part. B* **27**(2): 109-131.

ARR (1995). *Capacity and timing signals*.

Avineri, E., Prashker, J.N. (2003). The impact of travel time information on travelers' learning under uncertainty. *Proceedings of the 10th International conference on Travel Behavior Research*, Lucerne, Switzerland.

Avineri E., P., J.N. (2003). Sensitivity to Uncertainty: The Need for a Paradigm Shift. *Transportation Research Record* **1854**: 90-98.

AVV (1998). *Filekosten op het Nederlandse hoofdwegennet in 1997*. Rotterdam.

Barcelo, J. (2003). *AIMSUN User Manual*, V 4.1.3.

Barceló, J., Ferrer, J.L., Montero, L., Perarnau, J. (2004). Microscopic Simulation With Aimsun For The Assessment Of Incident Management Strategies. *European Transport Conference, ETC2004*, Strasbourg, France.

Bates, J., Polak, J., Jones, P., Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E* **37**: 191-229.

Bechara, A., Damasio, A.R. (2005). The somatic marker hypothesis: a neural theory of economic decision. *Games and Economic Behavior* **52**(2): 336-372.

Beckmann, M. J., McGuire, C.B., Winsten, C.B. (1956). *Studies in the economics of Transportation*. New Haven, Yale University Press.

Ben-Akiva, M., Lerman, S.R. (1985). *Discrete Choice Analysis*, MIT Press.

Benekohal, R. F., El-Zohairy, Y. M. (1999). Progression adjustment factors for uniform delay at signalized intersections. *Transportation Research Record* **1678**: 32-41.

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis (exposition of a new theory on the measurement of risk). *Commentarii Academiae Scentiarum Imperialis Petropolitanae* **5**: 175-192.

Bliemer, M. C. J. (2007). Dynamic queuing and spillback in an analytical multiclass dynamic network loading model. *Submitted for publication in Transportation Research Record*.

Bogers, E. A. I., Van Zuylen, H.J. (2004). The importance of reliability in route choices in freight transport for various actors on various levels. *European Transport Conference*, Strasbourg.

Bogers, E. A. I., Viti, F., Hoogendoorn, S.P. (2005). Joint Modeling of ATIS, habit and learning impacts on route choice by laboratory simulator experiments. *Transportation Research Record* **1926**: 189-197.

Bogers, E. A. I. (forthcoming). *The impact of traffic information to travel behavior*. Transport and planning. Delft, Delft University of Technology. PhD thesis.

Boillot, F., Blosseville, J.M. (1992). Optimal control of urban traffic networks. *6th International Conference on Road Traffic Monitoring and Control*, England.

Bonneson, J. A., Messer, C.J.and Fambro, D.B. (1988). Examination of shared lane operations. *67th Annual Meeting of the Transportation Research Board*, Washington D.C.

Bovy, P. H. L. (2001). Traffic flooding the Low Countries: how the dutch cope with motorway congestion. *Transport Reviews* **21**(1): 89-116.

BPR (1964). *Traffic Assignment Manual*. U.S DOT.

Branston, D., Van Zuylen. H.J. (1978). The estimation of saturation flow, effective green time and passenger car equivalents at traffic signals by multiple linear regression. *Transportation Research Part B* **12**: 47-53.

Brilon, W., Wu, N. (1990). Delays at fixed time signals under time-dependent traffic conditions. *Traffic Engineering and Control* **31**(12): 623-631.

Brilon, W., Geistefeldt, J., Regler, M. (2005). Reliability of freeway traffic flow: a stochastic concept of capacity. *16th International Symposium on Transportation and Traffic Theory*, University of Maryland, College Park, Maryland, Elsevier.

Brookes, D., Bell, M.G. (1991). Expected delays and stop calculation for discrete adaptive traffic signal control. *1st International Symposium on Highway Capacity*, Karlsruhe, Germany.

Burrow, I. J. (1987). *OSCADY: A computer program to model capacities, queues and delays at isolated traffic signal junctions*. TRRL Research Report. (UK).

Cantarella, G. E., Cascetta, E. (1992). Dynamic processes and equilibrium in transportation networks: Towards a unifying theory. *Transportation Science* **29**: 305-329.

Cascetta, E., Cantarella, G.E. (1991). A Day-to-Day and Within-Day Dynamic Stochastic Assignment Model. *Transportation Research Part A* **25**(5): 277-291.

Cascetta, E. (2001). *Transportation System Engineering: Theory and Methods*, Kluiwer Academic Publishers.

Catling, I. (1977). A time dependent approach to junction delays. *Traffic Engineering and Control* **18**: 520-526.

Chen, R. B., Mahmassani, H.S. (2004). Travel Time Perception and Learning Mechanisms in Traffic Networks. *In Proceedings of the 83rd TRB Annual Meeting*, Washington D.C.

Clark, S., Watling, D. (2005). Modelling network travel time reliability under stochastic demand. *Transportation Research Part B* **39**(2): 119-140.

Clayton, A. (1941). Road traffic calculations. *Journal of the Institute of Civil Engineering* **16**(7-8): 247-284,588-594.

Cotterill, P. J., Moore, S.E., Tudge, R. (1984). Intersection simulation model: INSECT. *12th ARRB Conference*.

Courage, K. G., Papapanou, P.P. (1977). Estimation of delay at traffic actuated signals. *Transportation Research Record* **630**: 17-21.

Cowan, R. (1978). An improved model for signalized intersections with vehicle actuated control. *Journal of Applied Probability* **15**: 384-396.

Cronje, W. B. (1983). Analysis of existing formulas for delay, overflow queues and stops. *Transportation Research Record* **905**: 89-93.

Daganzo, C. F. (1994). The Cell Transmission Model: a dynamic representation of highway traffic consistent with hydrodynamic theory. *Transportation Research Part B*, **29,** 269-287.

Davidson, K. B. (1966). A flow-travel time relationship for use in transportation planning. *Australian Road Research Bulletin* **8**(1): 32-35.

De Groot T., H. M. (2004). *Travel Simulator Laboratory (TSL)*. Delft, Delft University of Technology. Master thesis.

De Palma, A., Picard, N. (2005). Route choice decision under travel time uncertainty. *Transportation Research Part A* **39**(4): 295-324.

Department of Transport, U. K. (2000). *Road traffic statistics 2000*, U.K. Department of Transport.

Dion, F., Rakha, H., Kang, Y.-S. (2005). Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections. *Transportation Research Part B* **38**(2): 99-122.

Drew, D. R. (1965). Deterministic aspects of freeway operations and control. *Highway Research Record* **99**: 48-58.

Dunne, M. C. (1967). Traffic Delay at a signalized intersection with Binomial arrivals. *Transportation Science* **1**: 24-31.

Fambro, D. B., Rouphail, N.M. (1993). Generalized delay model for signalized intersections and arterial streets. *Transportation Research Record* **1572**: 112-121.

Fambro, D. B., Rouphail, N.M. (1997). Generalized Delay Model for Signalized Intersections and Arterial Streets. *Transportation Research Record* **1572**.

FHWA (1984). *TRANSYT-7F Users' Manual*, University of Florida.

FHWA (1999). *CORSIM User's Manual version 4.32*. McLean, Virginia, Office of Safety and Traffic Operations R&D, Intelligent Systems and Technology Division.

FHWA (2006). *Traffic analysis tools*. from http://ops.fhwa.dot.gov/trafficanalysistools/type_ tools.htm.

Fisk, C. S. (1988). Lane choice and intersection performance. *Urban road traffic models for economic appraisal stage III*, Volume **1** (Report No. 5, Appendix I), University of Auckland, department of Civil Engineering.

Friedrich, B. (2002). Adaptive signal control: an overview. *13th Mini Euro Conference – Handling Uncertainty in the Analysis of Traffic and Transportation systems*, Bari.

Fu, L., Hellinga, B. (2000). Delay variability at signalized intersections. *79th Annual Meeting of the Transportation Research Board*, Washington D.C.

Gartner, N. H. (1982). *Demand Responsive Decentralized Urban Traffic Control. Part I: Single Intersection Policies; Part II: Network Extensions*, Office of University Research, U.S. D.O.T.

Gerolominis, N., Skabardonis, A. (2005). Prediction of arrival profiles and queue lengths along arterials using a Markov decision process. *Proceedings of the 84th Annual Meeting of the Transportation Research Board*, Washington D.C.

Greenberg, H. (1959). An analysis of traffic flow. *Operations Research* **7**: 79-85.

Greenshields, B. D. (1935). The study in highway capacity. *Proceedings of the Highway Research Board*.

Haight, F. A. (1959) Overflow at traffic light. *Biometrika*, 46**,** 420-424.

Heidemann, D. (1994). Queue length and delay distributions at traffic signals. *Transportation Research Part B* **28**(5): 377-389.

Helbing, D. (1997). Modeling multilane traffic flow with queuing effects. *Physica A* **242**: 175-194.

Henry, J., Farges, J., Tuffal, J. (1983). The PRODYN Real-Time Traffic Algorithm. *4th IFAC-IFIC-IFORS on Control in Transportation Systems*, Baden-baden.

Hillier, F. S., Lieberman, G.J. (2001). *Introduction to operations research*. New York, U.S.A.

Hillier, J. A., Rothery, R. (1967). The synchronization of traffic signals for minimum delays. *Transportation Science* **1**(2): 81-94.

Holroyd, J., Hillier, J.A. (1971). *The Glasgow Experiment: PLIDENT and after*. Technical Report.

Hoogendoorn, S. P. *Travel Simulator Laboratory*. from www.tsl.tudelft.nl.

Hoogendoorn, S. P. (1996). *Multiclass continuum modeling of multilane traffic flow*. Transport and Planning. Delft, Delft University of Technology. PhD.

Horowitz, A. J. (1978). The subjective value of time spent in travel. *Transportation Research* **12**: 385-393.

Hunt, P. B., Robertson, D.I., Bretherton, R.D., Royle, M.C. (1982). The SCOOT online traffic signal optimization technique. *Traffic Engineering and Control* **23**: 190-192.

ITE (1995). *Canadian Capacity Guide for Signalized Intersections*. D. Institute of Transportation Engineers, Canada.

Jumsan, K., Rhee, S., Hwang, Z (2005). Vehicle passing behaviour through the stop line of signalised intersection. *Journal of the Eastern Asia Society for Transportation Studies* **6**: 1509-1517.

Kang, Y.-., S. (2000). *Delay, stop and queue estimation for uniform and random traffic arrivals at fixed-time signalized intersections*. Blacksburg, Virginia, Virginia Polytechnic Institute and State University. PhD thesis.

Kimber, R. H., Hollis, M. (1979). *Traffic queues and delay at road junctions*. TRL LR.

Law, A. M., Kelton, W.D. (2000). Simulation modeling and analysis. *Australian Road Research Bulletin* **3**.

Li, J., Rouphail, N. and Akcelik, R. (1994). Overflow delay estimation for intersections with fully actuated signal control. *73rd TRB Annual Meeting*, Washington D.C.

Lighthill, M. J., Whitham, G.B. (1955). On kinematic waves II: a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London*.

Lin, F.-B., Mazdeysa, F. (1983). Estimating average cycle lengths and green intervals of semi-actuated signal operations for level-of-service analysis. *Transportation Research Record* **905**: 33-37.

Liu, H., Lint, J.W.C. van, Van Zuylen, H.J., Viti, F. (2006). Urban travel time prediction based on queue estimation. *11th IFAC-symposium on Control in Transportation Systems*, Delft.

Lomax, T., Schrank, D., Turner, S., Margiotta, R. (2003). *Selecting travel reliability measure*s, Texas transportation institute Cambridge systematics Inc.

Lowrie, P. R. (1982). SCATS principles, methodologies, algorithm. *IEEE Conference on Road Traffic Signal*, England.

Luo, X., Lo, H.K. (2003). Travel time budget in transportation networks with stochastic degradations. *Proceedings of the 8th Conference of Hong Kong Society for Transportation Studies*, Hong Kong, China.

Markov, A. A. (1971). *Extension of the limit theorems of probability theory to a sum of variables connected in a chain*. Reprinted in Appendix B of: R. Howard. Dynamic Probabilistic Systems. Volume 1: Markov Chains.

May, A. D., Keller, H.E.M. (1967). A deterministic queuing model. *Transportation Research* **1**: 117-128.

Mayne, A. J. (1979). *An outline of new comprehensive formulas for queues and delays*, University College, London.

McNeill, D. R. (1968). A solution to the fixed cycle traffic light problem with compound Poisson arrivals. *Journal of Applied Probability* **5**: 624-635.

Meissl, P. (1963). Zufallsmodell einer Lichtsignalage mit mehrspurigem Stauraum. *Mathematik-Teknik-Wirtschaft*. Wien: 63-38.

Michalopoulos, P. G., Stephanopoulos, G. (1980). An application of shockwave theory to traffic signal control. *Transportation Research Part B* **15**(1): 35-51.

Michalopoulos, P. G., Pisharody, V.B. (1981). Derivation of delays based on improved macroscopic traffic models. *Transportation Research Part B* **15**(5): 299-317.

Miller, A. S. (1963). Settings for fixed-cycle traffic signals. *Operational Research Quarterly* **14**: 373-386.

Miller, A. S. (1968). The capacity of signalised intersections in Australia. *Australian Road Research Bulletin* **3**.

Minderhoud, M. M. (1999). *Supported driving: Impacts on motorway traffic flow*. Transport and Planning. Delft, Delft University of Technology. PhD Thesis.

Morris, R. W. T., Pak-Poy, P.G. (1967). Intersection control by vehicle actuated signals. *Traffic Engineering and Control* **10**: 288-293.

Mung, G. K. S., Poon, A.C.K., Lam, W.H.K. (1996) Distribution of queue lengths at fixed time traffic signals. *Transportation Research Part B*, **30,** 421-439.

Newell, G. F. (1965). Approximation methods for queues with application to fixed-cycle traffic light. *SIAM Review* **7**.

Newell, G. F. (1969). Properties of Vehicle Actuated Signals: I. One-Way Streets. *Transportation Science* **3**: 30-52.

Newell, G. F. (1971). *Applications of queuing theory*. Monographs of Applied Probability and Statistics, Chapman and Hall

Newell, G. F. (1990). Stochastic Delays on Signalised Arterial Highways. *11th international symposium on Transportation and Traffic Theory*. New York, Elsevier: 589-598.

Ngoduy, D., Hoogendoorn, S.P. (2004). Macroscopic traffic flow modeling and simulation of weaving sections on freeways. *Transportation Research Record* **1876**.

Ngoduy, D. (2006). *Macroscopic discontinuity modeling for multiclass multilane traffic flow operations*. Transport and Planning. Delft, Delft University of Technology. PhD.

Nicholson, A., Du, Z.-P. (1997). Degradable transportation systems: An integrated equilibrium model. *Transportation Research Part B* **31**(3): 209-223.

Ohno, K. (1978). Computational algorithm for a fixed cycle traffic signal and new approximate expressions for average delay. *Transportation Science* **3**: 99-125.

Olszewski, P. (1990). Modeling of Queue Probability Distribution at Traffic Signals. *Transportation and Traffic Theory*. M. Koshi. Amsterdam, Elsevier: 549-588.

Olszewski, P. (1990). Traffic Signal Delay for Non-uniform Arrivals. *Transportation Research Record* **1287**.

Olszewski, P. (1994). Modeling probability distribution of delay at signalized intersections. *Journal of Advanced Transportation* **28**(3): 253-274.

Ossen, S. J. L., Hoogendoorn, S.P., Gorte, B.G.H (2006). Inter-driver differences in car-following: a vehicle trajectory based study. *Transportation Research Record* in press.

Pacey, G. M. (1956). *The Progress of a bunch of vehicles released from a traffic signal*. London, Road Research Laboratory.

PTV (2003). *VISSIM version 3.70 user manual*

Pursula, M. (1999). Simulation of traffic systems - an overview. *Journal of Geographic Information and Decision Analysis* **3**(1): 1-8.

Ran, B., Boyce, D. (1996). *Modeling dynamic transportation networks*. Berlin, Springer-Verlag.

Richards, P. I. (1956). Shockwaves on the Highway. *Operations Research* **4**: 42-51.

Robertson D. I., Lucas, C.F., Baker, R.T. (1980) *Coordinating traffic signals to reduce fuel consumption*. TRRL Report - LR 934.

Rorbech, J. (1968). Determining the queue length of the approach lanes required at signal controlled intersections on through highways. *Transportation Research* **2**: 283-291.

Rose, G., Taylor, M.A.P., Tisato, P. (1989). Estimating travel time functions for urban roads: options and issues. *Transportation Planning and Technology* **14**(1): 63-82.

Ross, S. M. (1996). *Stochastic Processes*. New York, U.S.A.

Rouphail, N., Akcelik, R. (1989). Progression adjustment factors at signalized intersections. *Transportation Research Record* **1225**: 8-17.

Rouphail, N., Akcelik, R. (1992). Oversaturation delay estimates with consideration of peaking. *Transportation Research Record* **1365**: 71-81.

Rouphail, N., Tarko, A., Li, J. (2000). *Traffic flow at signalized intersections*, Lieu H. Revised Monograph of Traffic Flow Theory, update and expansion of the Transportation Research Board (TRB) special report 165, "Traffic Flow Theory", published in 1975.

Schrank, D., Lomax, T. (2005). *The 2005 Urban Mobility Report*, Texas Transportation Institute, The Texas A&M University System.

Sisiopiku, V. P., Rouphail, N.M. (1995). Toward the use of detector output for arterial link travel time estimation: a literature review. *Transportation Research Record* **1457**: 158-165.

Skabardonis, A., Dowling, R. (1997). Improved speed-flow relationships for planning applications. *Transportation Research Record* **1572**: 18-23.

Stephanopoulos, G., Michalopoulos, P.G. (1979). Modeling and analysis of traffic queue dynamics at signalized intersections. *Transportation Research Part A* **13**: 295-307.

Taale, H., Van Zuylen, H.J. (2003). Effects of anticipatory control with multiple user classes. *European Journal of Transport and Infrastructure Research* **3**(1): 61-76.

Tampere, C. M. J. (2004). *Human-kinetic multiclass traffic flow theory and modelling, with application to Advanced Driver Assistance Systems in congestion*. Transport and Planning. Delft, Delft University of Technology. PhD.

Tarko, A., Rouphail, N., Akcelik, R. (1993). Overflow delay at signalized intersection approach influenced by an upstream signal: an analytical investigation. *Transportation Research Record* **1398**: 82-89.

Tarko, A., Rouphail, N. (1995). A Distribution-Free Model for Estimating Random Queues in Signalised Networks. *Transportation Research Record* **1457**.

Tarko, A. (2000). Random Queues in Signalised Road Networks. *Transportation Science* **34**(4): 415-425.

Teply, S. (1989). Accuracy of delay surveys at signalized intersections. *Transportation Research Record* **1225**.

Thomas, T., Weijermars, W. (2006). Periodic Variations in Urban Traffic Flows *Seminar Infrastructure Reliability*. Delft.

Tian, Z. Z., Urbanik, T., Engelbrecht, R., Balke, K. (2002). Variations in capacity and delay estimates from microscopic traffic simulation models. *Transportation Research Record* **1802**: 23-31.

Tian, Z. Z., Wu, N. (2006). Probabilistic model for signalized intersection capacity with a short right-turn lane. *Journal of Transportation Engineering* **132**(3): 205-212.

Tijms, H. C. (2003). *A first course in stochastic models*. Padston, Cornwall, U.K., John Wiley & Sons, Inc.

Trafficware Corporation (1999). *Synchro 4, Traffic Signal Control Software*. Albany, California.

TRB (1994). *Highway Capacity Manual*, *special report 209*. N. R. Council. Washington D.C.

TRB (2000). *Highway Capacity Manual*. *Specia;l report 209.* Washington D.C., National Research Council.

Troutbeck, R. J., Brilon W. (2000). *Unsignalized intersection theory*, Lieu H. Revised Monograph of Traffic Flow Theory, update and expansion of the Transportation Research Board (TRB) special report 165, "Traffic Flow Theory", published in 1975.

Tu, H., van Lint, J.W.C., van Zuylen, H.J. (2006). Travel time variability versus freeway characteristics. *9th International IEEE Conference on Intelligent Transportation Systems*, Toronto, Canada.

Uchida, T., Iida, Y. (1993). Risk *assignment: a new traffic assignment model considering risk of travel time variation. Transportation and Traffic Theory: Proceedings of the 12th International Symposium*, Elsevier, Amsterdam.

USDOT (1996). *Traffic control systems handbook*. Federal Highway Administration.

Van Aerde, M. (1995). A single regime speed-flow-density relationship for congested and uncongested highways. *74th Annual Meeting of the Transportation Research Board*, Washington D.C., TRB.

Van Aerde, M. a. a. (2001). *INTEGRATION Release 2.30 - Users' guide*.

Van As, S. C. (1991). Overflow Delay at Signalised Networks. *Transportation Research Part A* **25**(1): 1-7.

Van Berkum, E., van der Mede P. (1992). *The impact of traffic information: Dynamics in route and departure time choice*. Delft, Delft University of Technology. PhD Thesis.

Van Lint, J. W. C. (2004). *Reliable travel time prediction for freeways*. TRAIL Press. Delft, Delft University of Technology. PhD.

Van Zuylen, H. J. (1985). De dynamiek van wachtrijen voor een verkeerslicht (The dynamics of queues at traffic lights). Dutch lecture notes VB41. Tilburg, Verkeersakademie.

Van Zuylen, H. J., Kikuchi, S. (2001). Traveler’s behaviour under uncertain conditions. *The Network Reliability of Transport*, M.G.H. Bell, Elsevier.

Van Zuylen, H. J., Taale, H. (2002). Dynamic and stochastic aspects of queues at signal controlled intersections. *13th Mini Euro Conference – Handling Uncertainty in the Analysis of Traffic and Transportation systems*, Bari.

Van Zuylen, H. J. (2003). *ITS for traffic signal control: design and evaluation*. Delft, Delft University of Technology, Course notes.

Van Zuylen, H. J., Viti, F. (2003). Uncertainty and the Dynamics of Queues at Signalized Intersections. *Proceedings CTS-IFAC conference 2003*, Tokyo, Elsevier.

Vincent, R. A., Pierce, J.R. (1988). *MOVA: Traffic Responsive Self-Optimizing Signal Control for Isolated Intersections*.

Viti, F., Van Zuylen, H.J. (2004). Modeling overflow queues on urban signalized intersections. *10th World Conference of Transport Research*, Istanbul, Turkey.

Viti, F., Van Zuylen, H.J. (2004). Modeling queues at signalized intersections. *Transportation Research Record* **1883**: 68-77.

Viti, F., Van Zuylen, H.J. (2005). Modeling overflow queues at signalized arterial corridors and with time-dependent controls. *84th Annual Meeting of the Transportation Research Board*, Washington D.C.

Viti, F., Van Zuylen, H.J. (2006). Concistency of random queuing models at signalized intersections. *85th Annual Meeting of the Transportation Research Board*, Washington D.C.

Von Neumann, J., Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, New Jersey, Prinseton University Press.

Webster, F. V. (1958). *Traffic signal settings*. Road Research Lab. London, Her Majesty Sationary Office.

Wu, N. (1990). *Wartezeit und Leistungfahigkeit von Lichtsignalangen unter Berucksichtigung von Instationaritat und Teilgebundenheit des Verkehrs* (In German). Lehrstuhl fur Verkehrswesen. Bochum, Ruhr-Universitat Bochum. PhD Thesis.

Yperman, I., Logghe, S., Tampere, C., Immers, B. (2006). The Multi-Commodity Link Transmission Model for Dynamic Network Loading. *85th TRB Conference.* Washington D.C.

Zijpp, N. J. v. d., Kirwan, K., Viti, F. (2002). *On the predictability of travel time on motorway networks*. R. a. v. innovatie. Delft, Delft University of Technology.

A

Markov Chains

A.1 Stochastic models

Stochastic models are strictly related to the definition of a stochastic process. A definition of stochastic process can be found in (Ross, 1996):

> A stochastic process $\mathbf{X} = \{X(t), t \in T\}$ is a collection of random variables. That is, for each t in the index set T, $X(t)$ is a random variable. We often interpret t as time and call $X(t)$ the state of the process at time t.

Therefore, stochastic processes deal with phenomena explainable with probability distributions and that are often correlated in time. A mathematical formulation, which explicitly explains the time dependency, is the following:

$$X(t) = E\{X(t)\} + \varepsilon(t) \quad \text{(A.1)}$$

The stochastic process $X(t)$ is then subdivided into two terms, a *forecasted value* $E\{X(t)\}$ plus a *forecasting error* $\varepsilon(t)$, where the error follows some probability distribution.

A.2 Poisson process

Among the stochastic processes, the Poisson process is a counting process that counts the number of occurrences of some specific event through time. Application of Poisson processes include the counting of arrivals at a service point, or the occurrence of some natural phenomenon (earthquake, flood etc.) in a certain area. Definition of Poisson process can be found in (Tijms, 2003):

> The counting process $\{N(t), t \geq 0\}$ is called a Poisson process with rate λ if the intercurrence times $X_1, X_2 ...$ have a common exponential probability distribution function:
>
> $$P\{X_n \leq x\} = 1 - e^{-\lambda x} \qquad \text{(A.2)}$$

Figure A.1 displays the distribution of arrival headways using Formula (A.2) for an average arrival rate of 0.1 vehicles per second.

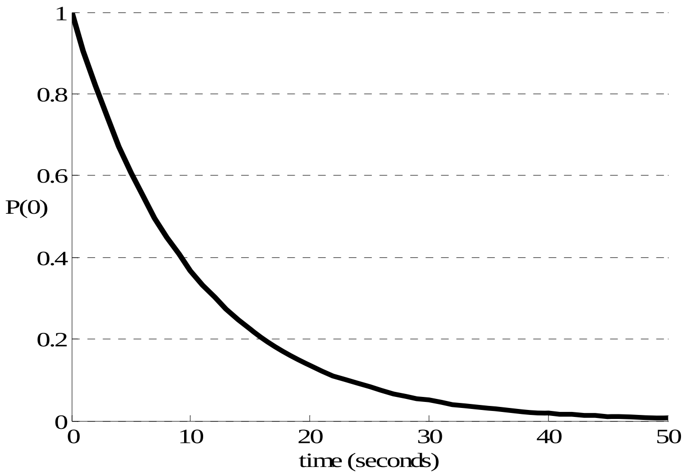

Figure A.1: Distribution of arrival headways for an average arrival rate of 0.1 veh/s

Poisson processes have been shown to well represent a large variety of real phenomena characterized by independent stochastic processes and each process having a very small probability of occurrence. The property of independence of random variables in a Poisson process implies that the k independent draws out of a Poisson process, with a common exponential distribution, have a probability of occurrence in an interval of time t that follows the Poisson probability function:

$$P\{N(t) = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \qquad \text{(A.3)}$$

A very important property of Poisson processes is the *memoryless property*, that is, at each point in time the waiting time until the next event has the same exponential distribution, regardless of how long ago the last event occurred. The memoryless property is a peculiar characteristic of Poisson processes and it justifies the mathematical tractability of such processes.

A.3 Renewal process theory

Many stochastic processes are regenerative, that is, they show the same probabilistic behavior from time to time. The time interval between two regenerations epochs is called a cycle. The sequence of regeneration cycles is commonly referred to as *renewal process*. Renewal theory is a generalization of the Poisson process, and it accordingly concerns the study of events occurring in sequence of time. Historically formulated to solve failure and replacement problems, renewal theory has been used for a wide variety of practical applications (among others, queuing systems, inventory, reliability).

The average of the random variable $N(t)$ is called the *renewal function*. The *excess variable* is the time elapsed from epoch t until the next renewal after epoch t and it is also called *residual life*. Renewal processes have, accordingly to the Poisson processes, the property of independence among variables, while the memoryless property is substituted by the cyclic property. This property gives some sort of dependency among the possible states of the stochastic variables. The easiest of the dependency is when the probability at one state depends only on the probability one step backward. In this case the renewal process is said to have a Markovian property (Markov, 1906) and the renewal process is said a *Markov Chain*.

A definition of Markov renewal process can be found in (Hillier, 2001):

> A stochastic process $\{Q_t\} = \{Q_0, Q_1, Q_2, ...\}$ is said to have *Markovian property* if:
>
> $$\Pr\{Q_{t+1} = j \mid Q_0 = k_0, Q_1 = k_1, ...Q_{t-1} = k_{t-1}, Q_t = i\} = \Pr\{Q_{t+1} = j \mid Q_t = i\}, \quad \forall \{i, j, k_0, k_1, k_2, ..., k_{t-1}\} \tag{A.4}$$
>
> The conditional probability of any future state, given any past state and present state, is independent of the past states and depends only on the present state.

Markov chain processes under this hypothesis are defined as one-step transition models. This property gives the opportunity to model dynamic stochastic processes characterized by non-stable behavior (i.e. day-to-day learning processes) in a very simple way. If the following condition holds:

$$\Pr\{Q_{k+1} = j \mid Q_k = i\} = \Pr\{Q_{t+1} = j \mid Q_t = i\}, \quad k \in [0,t], \quad \forall\{i, j, t > k\} \tag{A.5}$$

then the Markov process is said to be *stationary*, that is the probability distribution does not change in time. A condition that guarantees a state to be *transient* is also given in (Hillier, 2001):

> A state is said to be *transient* if, upon entering this state, the process *may* never return to this state again.

Consequently, a transient state can be visited only once. Another property relevant in this thesis is the absorbing property, which is typical from states that gradually tend to steady state:

> A state is said to be an *absorbing* state if, upon entering this state, the process *never* will leave this state again.

Conditional probabilities can be represented in matrix form. Using the notation $p_{ij}(t) = \Pr\{Q_{t+1} = j \mid Q_t = i\}$ an n-step transition matrix is formulated as the following:

$$\mathbf{p}(t) = \begin{bmatrix} \mathrm{p}_{00}(\mathrm{t}) & \mathrm{p}_{01}(\mathrm{t}) & \dots & \mathrm{p}_{0N}(\mathrm{t}) \\ \mathrm{p}_{10}(\mathrm{t}) & \mathrm{p}_{11}(\mathrm{t}) & \dots & \mathrm{p}_{1N}(\mathrm{t}) \\ \dots & \dots & \dots & \dots \\ \mathrm{p}_{N0}(\mathrm{t}) & \mathrm{p}_{N1}(\mathrm{t}) & \dots & \mathrm{p}_{NN}(\mathrm{t}) \end{bmatrix} \tag{A.6}$$

Each element represents the probability of observing an evolution from the state in row to the state in column. The matrix has to satisfy the condition:

$$\sum_{j=0}^{N} p_{ij}(t) = 1 \tag{A.7}$$

If transition probabilities do not depend on t the transition matrix is then stationary. The transition probability, in combination with the probability distribution of the previous cycle, gives the probability of a state i to occur at time t as:

$$p_i(t) = \sum_{j=0}^{N} p_{ij}(t) \cdot p_i(t-1) \tag{A.8}$$

To solve this equation it is then simply required the specification of some initial conditions $p_i(0)$ for every state i.

B

The travelers' response to uncertainty

B.1 Introduction

Assessing the influence of travel time variability on travelers' decisions has been addressed as one of the challenges in recent transportation research. Several authors emphasized the importance of including uncertainty about travel conditions as a factor influencing travel decisions, especially in terms of route choice and departure time choice (i.e. (Avineri and Prashker, 2003), (Chen and Mahmassani, 2004), (de Palma and Picard, 2004)). This appendix presents empirical evidence of how much people value the variability of alternatives in comparison with the expected values, in relation with the experience and the information they have at the decision time.

When people face a decision that yields an uncertain result, their decision is sometimes not easy or straightforward. Variability of costs, conflicting objectives, competing alternatives and heterogeneous risk attitudes make decision making somewhat variable among travelers and consequently difficult to predict. High variability of travel times limits the possibility for users to properly estimate the benefits associated to an alternative. Moreover, memory limits human minds and impedes a correct estimation of an average value. For example, people tend to forget or exaggerate unhappy events of the past (Van Zuylen and Kikuchi, 2001) deviating their perceptions from objective reality.

Behavioral models assume travelers to choose their preferred routes according to the costs they expect for all known alternatives. Conventional Random Utility Models (RUM) assume the users to have perfect knowledge of past and actual conditions and choose accordingly the alternative, which gives them the highest utility. The costs determine the systematic component of the utility while the perception of costs is usually randomized by including an error term. This term is introduced to model a noise factor to model a certain degree of perception error. In reality decisions may be made under complete or partial information about the real travel costs, or under time constraints, or even depending on the traveler's emotional state at the time the decision is made.

Travelers get to know about the expected costs of all available alternatives of travel from two sources: from their own past experiences, thus how many times they traveled using that alternative or another, and from the information they acquire from external sources, i.e. variable message signs, road maps, radio etc. Past experiences and information are combined in order to have a higher degree of confidence on the expected travel costs (Bogers et al., 2005). When conditions are uncertain travelers may prefer reliable routes instead of risking a travel on an average shorter but uncertain one.

Experienced travel times and traveler's attitudes at each trip (his level of habit, curiosity, risk acceptance, experience and reliability towards roads and information) are combined to get an expectation of the utility the traveler might have from the next trip, and how uncertain these costs can be. Given the road conditions experienced during past trips and the ones perceived by external information, the users select the road and the departure time, which should maximize their utility taking into account the variability of this utility and the risk attitude they have.

In conclusion, the expected utility travelers may associate to each alternative is assumed to depend not only on the expected network costs and their variability, but also on the learning process that leads to this value, on the property of each individual to cope with uncertainty and with information they get before and during the trip. This decision-making process is assumed to vary with:

1) the *experience* a traveler has got regarding the available routes;

2) the *information* collected before and during the trip and

3) the *individual characteristics* of travelers, like habit, risk attitude, anxiety, motivation, etc.

The following considerations discuss the role of these three characteristics in the travelers' decisions in terms of route and departure time choice.

B.2 Experience

People get to know about the possible costs they may face by experiencing the routes for various days, and sometimes during different periods of the day. If a traveler has little knowledge about the status of the roads at the time of departure, he might assign a very large safety margin to the expected travel time, in order to avoid that large travel times cause also very large delays at the arrival point. If instead the traveler has experienced the roads quite often, given his past experiences he may exclude that at the time of departure delays or large travel times may occur. The level of experience may thus considerably influence the perception of uncertainty and the choice under risk.

The level of confidence given by the experience is then an influencing factor in the travelers' decision-making process. This level of confidence may be influenced by the number of travels experienced using an alternative, related to the variability of travel times, but also on travelers' individual characteristics like habit, curiosity, stress etc.

Travel time variability and experience are thus highly correlated and this relationship can affect the way flows reach equilibrium. Based on former route and departure time choices, the travelers have personal experiences. From these experiences, they can learn about the characteristics of the routes they have chosen, about how to interpret travel information and about the reliability of this information.

B.3 Information

Experience is usually combined with other sources of information in order to get a more accurate scenario of which costs a traveler may encounter until arrival at destination. While past experiences might help the traveler at having a guess of what kind of conditions might be encountered at a certain route and time and how much a trip "usually" costs, information can help at clarifying the actual status of the roads.

Information can give estimates of actual or future conditions. It is not possible to guarantee a perfect estimation of costs, even with the most accurate models, since information is subject to the variability and the dynamics of the traffic states. Travel time variability affects in fact the quality of information and the impact of the latter to the users. Some users may still rely on their personal opinion instead of following the suggestions eventually given by the information if information is frequently imprecise. Knowing the impact of travel time uncertainty when travelers combine past experiences with information is thus important in the assessment of ATIS systems.

B.4 Individual characteristics

Even if travelers increase their level of experience and information in order to reduce as much as possible their uncertainty about the expected costs of a trip, uncertainty still affects their choice if travel times are variable. Risk attitude of people can depend on various factors: on road conditions, trip purpose, traveler's stress level, individual's attitude, level of optimism and so forth. People tend to have different objectives and value differently the elements determining their utility. They may value in a different way driving times, or delays, or waiting times at queues. People may rely more on personal past experiences than information showing strong habit and sometimes it is the other way around.

This role of uncertainty in choice behavior is illustrated in the next section, which provides empirical results of the choice process of travelers under uncertain conditions in terms of route and departure time. The value of past experiences and information in their decision-making process is analyzed in a simple scenario involving two alternative routes and an interval of possible departure times. The way these characteristics are evaluated in the decision-making process of the travelers is analyzed using an internet survey. The reliability is shown to be a characteristic that clearly influences the preference of users towards certain routes. The data used to obtain these results have been collected with an SP (Stated Preference) experiment using the Travel Simulator Laboratory (Hoogendoorn, 2004).

B.5 Empirical findings

This section shows some results taken from two laboratory experiments made with the web-based tool TSL, designed at the Delft University of Technology (Delft, the Netherlands). People were asked to repeat many times a certain trip with a fixed origin and destination on the motorway network around the city of Amsterdam (see figure 2.6) and to select at each round a time for starting the trip and, after receiving the information about the expected status of the roads, the route to use.

Two major difficulties affect the choice analysis and valuation of the parameters: 1) data acquisition and 2) utility model selection. The first problem arises since information about travel choices and decisional variables is hardly complete; if travel choices are observed from reality, it is difficult to catch what have determined these choices since decisional variables are hardly controllable in real life. On the other hand, if choices are determined by simulating different scenarios and asking people's preferences, a better knowledge and control of the variables is achieved, but question is whether a traveler would act similarly in real life; the result is then highly dependent on the way trips are simulated. The second problem is related to the form and the complexity of the model selected to represent the utility of a traveler; very simple models can poorly fit the results of the data, or miss some important characteristic that determine the travelers' choice. On the other hand, very detailed and advanced models have the risk of overfitting the data and therefore give results that are hardly applicable to other data than the calibrated. For a more detailed description of the problems and limitations of discrete choice analysis one can refer to (Ben-Akiva 1985) or (Cascetta 2001).

B.5.1 Description of the experiments

First step of the experiment is the collection of some individual characteristics of the respondents, age, gender, driving experience etc. Later it was asked what their preferences were regarding the following travel time attributes: arriving early, in-vehicle travel time, arriving late. During the experiment, a score could be determined by the sum of the normalized weight for a route attribute times the number of minutes spent for that attribute.

Let α_{ik}, β_{ik}, γ_{ik} be respectively the weight assigned for a minute lost for early arrival, driving time and late arrival at destination, and $t_{ik}^{early}(n)$, $t_{ik}^{driving}(n)$, $t_{ik}^{late}(n)$ the respective quantity of minutes that respondent i loses selecting an alternative (route and departure time) k during step n. The score is computed for each individual by:

$$\mathrm{I}_{ik}(n) = \alpha_{ik} \cdot t_{ik}^{early}(n) + \beta_{ik} \cdot t_{ik}^{driving}(n) + \gamma_{ik} \cdot t_{ik}^{late}(n) \quad \text{(B.1)}$$

They were told that their objective was to minimize the sum of the scores over 25 rounds.

Table B.1: Characteristics (in minutes) of the routes in both experiments

		Experiment 1	Experiment 2
Route 1	Average	17.68	23.05
	Standard deviation	4.25	6.65
Route 2	Average	19.86	25.31
	Standard deviation	3.92	2.43

Travel times are calculated in TSL using a stochastic simulation model, which computes travel times according to a normally distributed loading demand and a stochastic network-loading model. Information is given, whenever the panel is on, displaying the length of queue in kilometers. To simulate the stress of waiting in a queue a waiting time proportional to the total travel time spent and computed with a speeding factor was applied.

The analysis of the two experiments is finalized to:

1) investigate the sensitivity of the users to information when routes are equally variable and when one route is clearly more reliable than the other, and quantify the relative importance of past experiences and information in the utility of the travelers;

2) show the relevant role of travel time reliability in the utility of travelers, both in departure time and in route choices;

3) analyze the influence of reliability in the learning process of users towards the individual preferred route and departure time with and without a reliable route;

The first experiment presents the choice of two routes with similar expected travel times and variance while the second scenario involves two routes where the travel time of the first is highly variable while the second has a sensibly higher expected travel time but very high reliability (see table B.1). For the first experiment data from 52 respondents were available; for the second one there were 63 respondents. It concerned highly educated people mainly from The Netherlands, Italy and Portugal, most of them being engineers. For a more detailed explanation about set-up of the scenarios and the design of the experiments see (De Groot and Hellendoorn, 2004).

B.5.2 Results

This section presents some results obtained from the analysis of the average response of users in terms of route and departure time and using the driving time, the lost time at arrival and the scores as performance measures.

The respondents' utility function for choosing the alternative k has been assumed linear and having a mixed Logit model structure, following the relationship:

$$U_{ik}(t) = \boldsymbol{\beta} \cdot \mathbf{X}_{ik}(t) + ASC_{ik} + \varepsilon \qquad \text{(B.2)}$$

The indexes i, k, t represent as the above formula (B.2) respectively the alternative index, the respondent index and the time step. $\boldsymbol{\beta}$ and $\mathbf{X}_{ik}(t)$ are respectively the vector of free parameters and the vector of explanatory variables related to person i and alternative k, which are found significant in the choice of users. ASC_{ik} represents the alternative specific constant related to person i and route k while ε is the error term, which models some choice variability. The choice of a mixed Logit is justified by the time-dependent characteristic of the choices for each respondent and the strict correlation among these choices. For a detailed description of the model and the model results see (Bogers 2005).

Looking at Figure B.1 one can have an insight into the learning mechanism of travelers. It appears that while the average travel time is not changing with the experience of the users in both experiments, the average lost time clearly decreases in time. A learning curve is clearly shown in Figure B.1 where a normalized score index is compared in time between the two experiments.

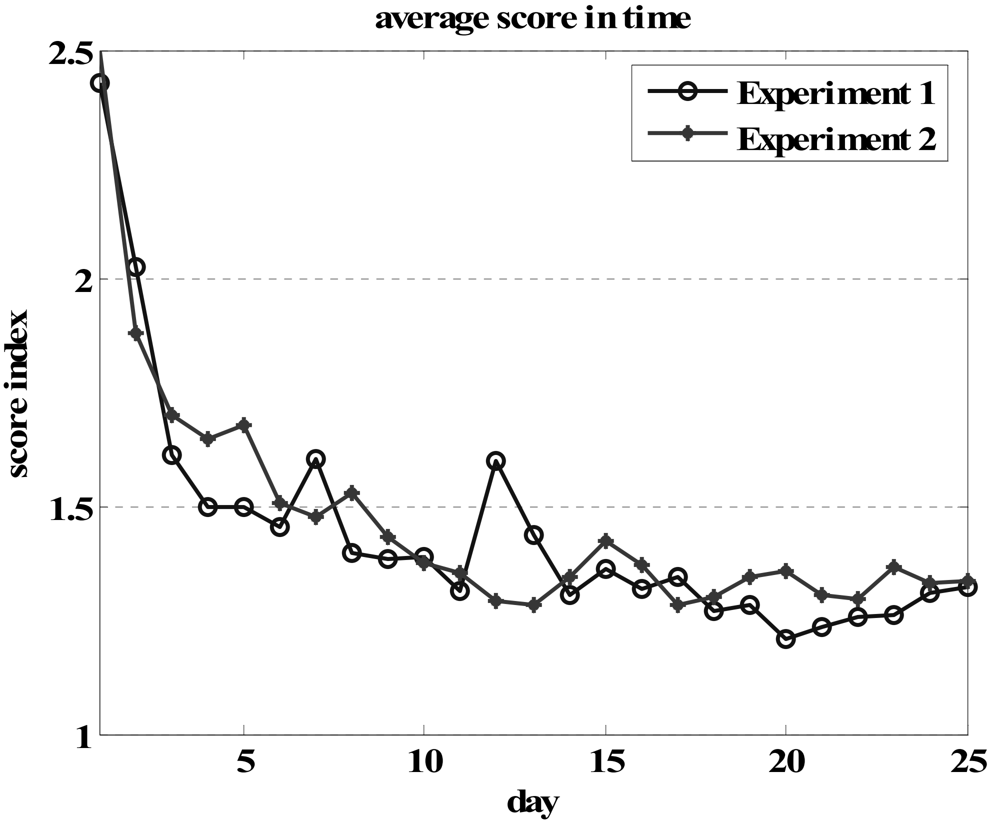

Figure B. 1 : day-to-day average evolution of the score index

The index is weighted by dividing each individual's score by the minimum travel time which the respondent could have experienced during the same day by selecting the proper route and/or departure time:

$$\text{score index} = I_{ik} / \min_{k}(I_{ik}) \qquad \text{(B.3)}$$

Interestingly, the presence of a reliable route in the system does not appear to change consistently the learning behavior of the respondent. In both experiments the average shows to gradually become less steep after about 10 days of gradual learning. Yet, the equilibrium is still far from the optimal choices (around 30% error). In conclusion a network system where travel times are affected by some uncertainty, travelers are not likely to perfectly choose the best actual alternatives but they tend to make some estimation and prediction errors. Equilibrium conditions can be very different if one takes into account this uncertainty in travelers' rational choice. For example, risk attitude of people in relation with the uncertainty in the network is a factor, which can contribute to having a better knowledge of travelers' tastes and give some reason for selecting non-optimal choices. Several studies in transportation deal with the influence of risk attitude in decision-making (e.g. (Avineri E. 2003), (E. A. I. Bogers, Viti, F., Hoogendoorn, S.P. 2005)). Bogers and Van Zuylen (E. A. I. Bogers, Van Zuylen, H.J. 2004) found from a survey involving truck drivers that individual risk aversion is in linear relationship with the relative variance of travel times between two route alternatives. An interesting direction of analysis is to evaluate then how travelers modify their risk aversion when receiving information about the status of the roads. In figure B.2 the fraction of respondents that in the second experiment decided to use route 2 is compared with the fraction of cases where route 2 was actually the shortest alternative.

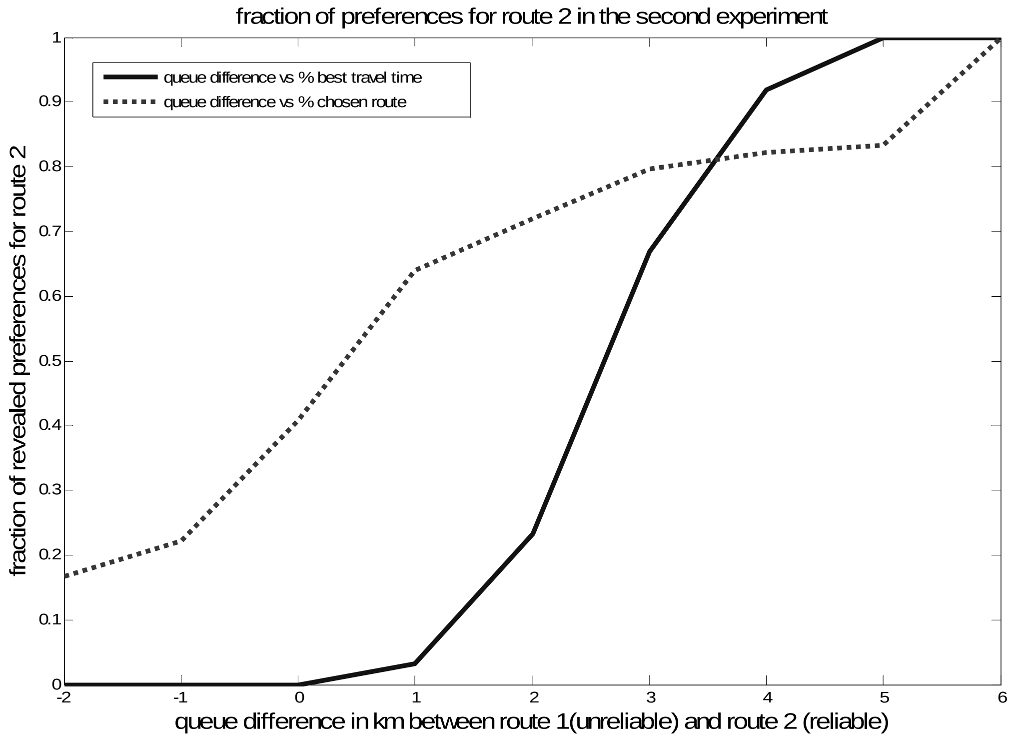

Figure B. 2 : Violation of the conventional Expected Utility from a repeated experiment

The curve of the best choices shows how information has been modeled as stochastic variable, and the curve is the representation of its cumulative distribution. The curve obtained by averaging the choice revealed by the respondents does not appear to follow the same behavior. Until route 1 has 25% of chances to be the shortest route the respondents overestimate the preference on route 2 showing risk aversion, while strangely it shows the inverse trend when route 1 has a lower probability to be the best alternative and some people become risk prone. In these results there is some evidence of the certainty effect and the inflating of small probability already found in Avineri and Prashker (Avineri 2003). The intersection point of the two curves drawn can be seen as the certainty equivalent related to the information, that is the users are indifferent between choosing the unreliable, or the on average longer but reliable alternative in those conditions. The underestimation of the preference in route 2 when it is clearly the best option can point at the second effect, thus some people gamble on the unreliable route feeling lucky.

The above experiments have shown empirical evidence of the relevant role of uncertainty and reliability of travel times in the decision-making of travelers. Both route choice and departure time choice are in some way affected by uncertainty in various ways. As seen in the last section, reliability appears to give positive utility to an alternative of travel.

C

Lane changing behavior in multilane sections

C.1 Introduction

Signalized intersections with multiple lanes per road section can be designed to dedicate one or more lanes to each specific traffic stream. Figure C.1 shows an example of how an intersection can be designed schematically assigning multiple lanes per traffic stream. In these situations the flow distribution among lanes can be very different. Lane changes upstream the intersection can in fact be influenced by a queue building downstream. Travelers might expect to have a smaller delay if they queue up behind the smallest queue. This maneuver can be on the other hand limited by the number of lane changes the vehicle should do to reach a smaller queue than the one the vehicle will find if keeping on driving on the same lane, and by the presence of other vehicles on the target and the intermediate lanes. Unbalanced distribution of flows among lanes can lead to an unbalanced distribution of queue lengths. On the other hand unbalanced distribution of queues can affect the lane changing behavior.

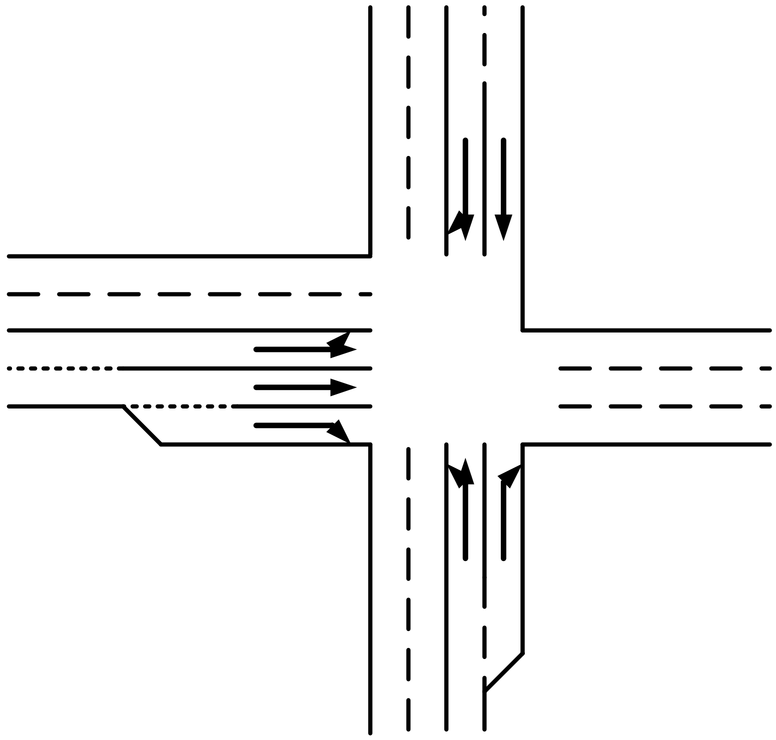

Figure C.1: example of intersection with multiple lanes

Most of models in literature deal with multiple lane sections as disaggregate, dealing with flows separately. The Highway Capacity Manual (TRB 2000) suggests some rules to define when some traffic streams should be separated or they should share a lane. Flows are considered equally distributed among lanes for the same flow stream if no estimation of flow distributions is available. As alternative, estimation of the expected distribution can be done for example by field data analysis.

C.2 Microscopic simulation of lane flows

Microscopic programs frequently are used in the evaluation of queues and delays at multilane intersections, given their property of treating each vehicle as a physical entity and the possibility to simulate lane changing at the individual vehicle level. VISSIM uses different parameters in the estimation of lane changing behavior approaching the signal than within a road section. Vehicles are assumed to have a smaller gap acceptance and an increased "degree of attention", thus lane changes are more frequent in these conditions. A lane change occurs in the program when the vehicle approaching the signal has still some distance to drive if changing to the adjacent lane. This implies that lane changes due to unbalanced queue in between lanes occur only at the moment vehicles reach the queue. In reality, frequent lane changes are observed in, for example, large toll plazas already some distance before the position of the back of the queue. Travelers show in this case an anticipatory behavior, changing lane depending on the queue they expect to find when they arrive at the intersection. The microscopic program AIMSUN (Barcelo 2003) shows some anticipatory behavior in this sense by applying a different lane changing rule at intersections. Lane changes are driven all throughout the network by the smallest path costs. If changing lane implies a reduction of the path cost, vehicles will try and change lane. Path costs are constantly updated using the average density, so that if queue builds up at one signal, vehicles will try and go to the queue that will imply the smallest route delay. Using this methodology, AIMSUN shows somewhat an anticipatory behavior of drivers approaching the intersection. Lane changes are also calibrated using a threshold value for the lane change decision, thus only a certain gain in terms of travel time reduction leads to a lane change desire. Furthermore, lane changes are balanced by a right-lane-driving rule.

To evaluate the lane changing behavior of the users crossing an intersection, a similar simple scenario, like the one displayed in figure 6.2, is evaluated using AIMSUN. The new scenario consists of two lanes with permitted lane changing all along the road section. Green time is set to again 24s and cycle is set to 60s. Different stationary demand conditions have been simulated, together with different split rates in between lanes. Detectors have been placed every 500m until the signal and one detector has been placed at each lane after the signal. To compute average and standard deviation 100 random simulations were done. Results are presented in the following of the section.

Figure C.2 shows the average flow rates detected on the two lanes. Scenarios 1 and 2 show a total demand of 1500veh/h and an initial split rate of respectively 15-85% and of 30-70%.

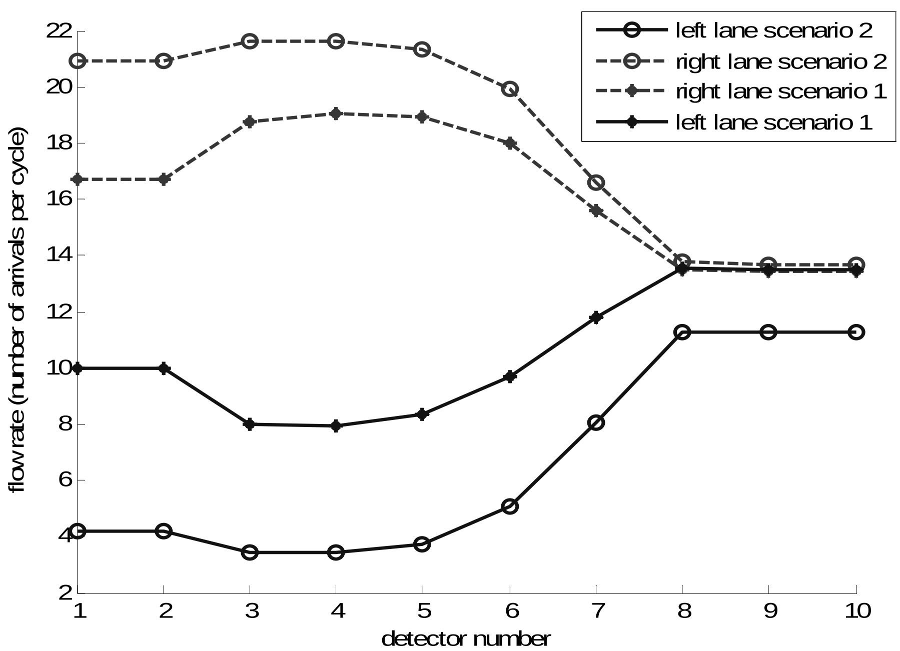

Figure C.2: Dynamics of the average flow distribution among lanes

When vehicles are loaded, a certain percentage moves to the right lane according to the right-lane-driving rule. The predictive property of vehicles approaching the intersection is shown after the 5th cycle, when split rates start to decrease in difference. Just before the intersection (detectors 8-9) the vehicles distribute nearly equally among lanes. Lane changes are only slightly influenced by the gap-acceptance model, since the assigned demand is near capacity, but far below the saturation flow.

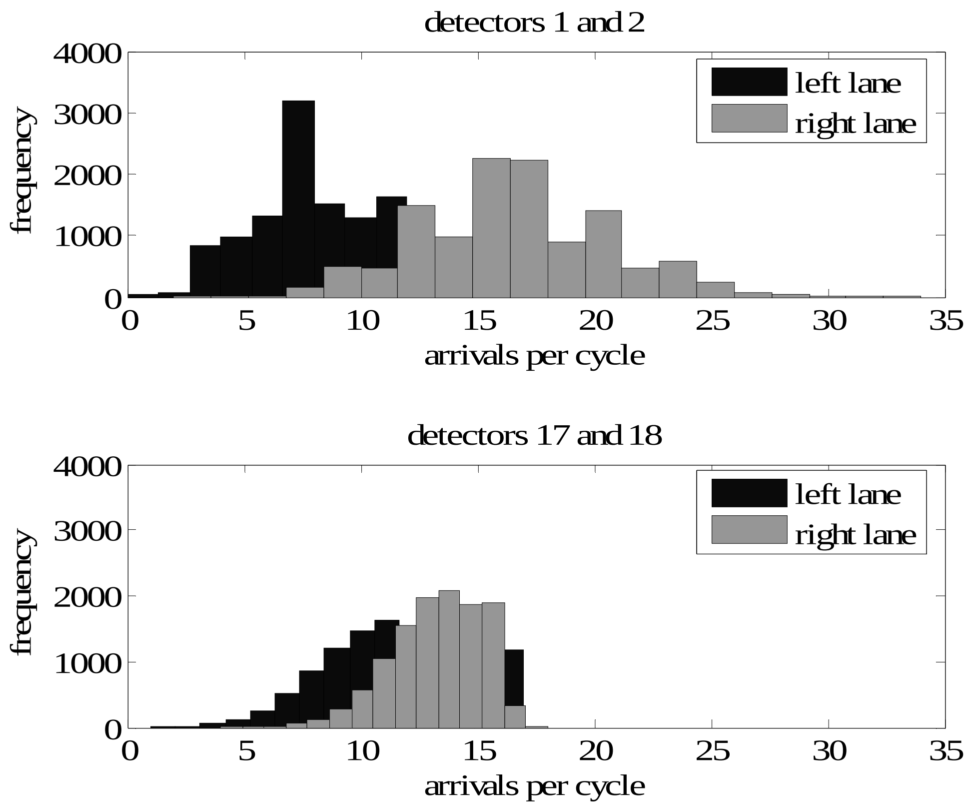

Figure C.3: histogram of vehicle counts at the start of the section and before the signal

Although the simplification made with the Highway Capacity Manual appears in these conditions correct, there are several reasons that make the hypothesis of equally distributed arrivals at intersections appear somewhat restrictive. Figure C.3 shows the distribution of vehicle counts from the detectors placed immediately after the origin of the road section and from the ones placed before the signal. A first interesting result is that the distributions of arrivals do not show very large outcomes. Although over 30 vehicles can be recorded in a cycle from the right lane and at the origin, very few times the detector before the signal has recorded more than 17 vehicles. This effect influences the variability of vehicle arrivals. In fact, if one computes the sum of the standard deviation of each lane, the variability of the system is reduced from around 7veh/cycle to 4.5veh/cycle. In conclusion, the presence of more than one lane for a traffic stream reduces the variability of arrivals, and consequently a reduction of the expectation value and the variability of the overflow queue should be expected too.

Estimating the queue length at multiple lane sections can also be useful in design problems, for example to evaluate the optimal length of exclusive turning lanes, taking into account spillback effects and the physical space available to build the lane.

About the author

Francesco Viti was born in Matera (Italy) in 1975. He lived in several cities in Italy with his family until he moved to Napoli in 1990, where he started his studies in the Transportation area at the University “Federico II” in 1994. He graduated with Honors in Civil Engineering, majoring in Transportation Engineering, doing a Master thesis which focuses on a model for the optimization of urban accessibility using a park pricing strategy. During his last year he attended a summer course at the MIT of Boston, the URIT2000, where he met his current supervisor, Prof. Henk van Zuylen.

In 2001 he joined the Dynamic Traffic Management group of Transport and Planning at the Delft University of Technology, starting a Ph.D. research which ended with the writing of this dissertation. During his stay at the TU Delft he focused his research mainly on the traffic flow operations at signalized intersections, producing several papers and presenting his work at various international conferences. He was awarded in July 2004 with the prize “Best Paper for a Young Researcher” at the 10th WCTR Conference in Istanbul.

Apart from his main research duties, he was often interested in other projects and extra activities. Among others, he joined the TRAIL PhD Council in 2003 and he became chairman of this council from January 2005. He was course leader of the course “Game Theory: theory and applications in transportation” given at the TRAIL Research School. He was also selected to represent the Transport and Planning section at the Assessment of Research Quality 2000-2004 of Civil Engineering at the Delft Univertity of Technology in June 2005 and at the TRAIL Research Assessment in September 2006. Moreover he gave his contribution to Ph.D. and project researches on travel information systems, pricing, travel time prediction, Advanced Driver Assistance Systems etc.

Summary

Traffic congestion levels at urban networks are becoming more and more severe and peaks of the demand are extending over longer time periods. A way to reduce this problem is adapting the infrastructure, e.g. by adding new lanes or building new roads. As an alternative, transport managers can improve the network conditions by using the available network infrastructure more efficiently. Some of these alternative management strategies are referred to as Dynamic Traffic Management (DTM) measures. DTM has the objective to improve traffic safety and the utilization of the transport infrastructure. This is obtained by dynamically adapting the available capacity of the infrastructure to the demand (e.g. by using adaptive signal controls) or vice versa (e.g. by using dynamic pricing). The optimization of these DTM measures needs traffic models to predict the impact of measures and to find the measures that optimize the performance of the traffic system. Among others, the increase of reliability is an important objective, related to safety and capacity. Therefore, the knowledge of the variability of the performance measures, e.g. travel times, delays etc. becomes important too.

The knowledge of travel time variability is fundamental in the evaluation of DTM strategies but also in travelers' choices. In fact, travelers take in consideration travel time variability sometimes even more than mean travel times. Especially when time constraints are involved (e.g. important meetings) a road user is very concerned about the risk of long unexpected travel times and is willing to spend extra time on the road if this is more predictable and implies less risk of late arrival. The importance given nowadays by road analyst and governments to the travel time reliability and the reduction of uncertainties, and the impact of travel time variability on road drivers should motivate as much research on these characteristics as for the expectation value of travel times.

Travel times at urban networks are for a large part determined by the delay drivers experience at (controlled) intersections; therefore, this thesis focuses primarily on signal control delays and their distribution, although the methodologies adopted to develop the models can be applied to several other operational and behavioral issues.

The scientific forum agrees that large improvements are still needed in the modeling of delays, above all, because of the variable behavior of queues forming and dissipating within a cycle and cycle-by-cycle. The queuing process at signalized intersections has been studied already for half a century but it still remains a weak point in the traffic flow

theory at urban networks. The long list of heuristic models developed in the past testifies that a comprehensive theory, which explains how these systems operate, is still missing.

Given the probabilistic nature of traffic, large uncertainty surrounds the estimation and prediction of the performance of signalized systems. Straightforwardly, only probabilistic traffic flow models are able to measure this uncertainty and to represent the dynamics of queues and delays without neglecting the long-term effects of the variability of traffic.

Probabilistic models use true macroscopic relationships between state, control variables and the resulting performance measures, and they assume these variables as statistically distributed according to a known probability distribution function. Consequently, also the performance measures are calculated in a probabilistic fashion. A probabilistic approach gives the opportunity to analyze the statistical properties of traffic and give estimates of the expected conditions and of the variability of traffic via the computation of e.g. standard deviation or 10-90% confidence values.

Starting from an earlier developed model of overflow queues at fixed time controls based on the renewal theory of Markov Chains, a probabilistic model that describes the queuing process also within each cycle was developed. This model allows the computation of the queue length at any time of each cycle; it extends the previously developed Markov model, which computes the queue length distribution using discrete time steps, and it justifies the dynamic and stochastic character of the overflow queues. The Markov Chain process presented in Chapter 4 represents a very powerful technique for modeling dynamic and stochastic processes, since it treats variables at the probabilistic level and it simulates traffic by exact expressions based on mass-balance equations. This computing property makes these models more suitable than microscopic models for planning purposes, since they simulate the variability of traffic and analyze causes-effect relationships between state variables, control variables and performance measures within reasonable computing times.

The newly developed model is derived without any assumptions about the statistical properties of the arrival process. It fits for Poisson, binomial, Normal distributed arrivals. More important than the possibility to calculate the random delay with a more general model than the existing ones is the insight that the derivation gives into the process that causes the random delay. The assumption used by some authors that the queue should be represented by a step function appears to be superfluous. The stepwise character of the delay is transformed to a smooth character of the expected queue length, linearly increasing in the red-phase and the first part of the green phase. The expectation value of the queue in the green phase shows a non linear character as soon as the tail of the probability distribution comes close to zero. This phenomenon causes the overflow delay.

This method has quantified the error that affects standard analytical procedures like the HCM 2000 to compute vehicle delays especially when the model should compute decreasing queues after an oversaturated period, which frequently occurs during peak periods. All available methods up to date underestimate vehicle delays, or flatten delays within an evaluation period, instead of computing more accurately delays for each cycle. Combined analysis of average and standard deviation of the queue in time shows strong interdependence among these two characteristics, especially in saturated conditions of traffic. Such conditions of traffic create a delay that propagates in time and causes extra waiting times for vehicles approaching the intersection even several cycles after congestion had occurred.

An exponential behavior has been found when queues recover from large initial values and the signal operates near capacity. This finding has suggested a new formulation for the dynamics of the overflow queue, called the Van Zuylen-Viti formula, which combines the deterministic linear behavior with a smoother asymptotic behavior towards the equilibrium value. Moreover, an expression for the standard deviation of the overflow queue is also proposed. The models proposed can compute queues and delays assuming both stationary and non-stationary demand conditions. This feature makes them suitable e.g. for Dynamic Loading processes, but also for model-based predictions of expected waiting times and for planning and design problems.

The Markov and the Van Zuylen-Viti formulas have been compared with the results of the VISSIM microsimulation software program showing very good agreement. The consistency between the three approaches in various conditions of traffic validates the two less detailed methods. This represents also an important contribution to traffic managers and practitioners, since it proves that the dynamics of the overflow queue are well estimated with all three different level-of-detail models.

The probabilistic approach allows one to model queues also under broader assumptions than the simple fixed time, single lane, and isolated intersection. While there is very little difference in the formulation of the Markov model for isolated intersections from an intersection within a network, i.e. the shape of the arrival distribution, modeling the interactions between lane choice of drivers and queue lengths appears more complex. To account for this interaction the Markov model has been combined with a probabilistic lane-changing model. By doing so, the distribution of arrivals has been shown to have a dynamic character, according to the dynamic character of the overflow queue length. Furthermore, the Markov model at multilane sections allows one to account for spillback effects, which is useful information for a correct estimation of delays and for the design of exclusive turning lanes. Finally the assumption of fixed control settings has been relaxed by formulating a probabilistic model of vehicle actuated controls. This approach allows one to compute the probability of green time extension depending on the variability of arrivals and their headway distribution in time.

The newly developed models presented in this thesis give important contributions to several areas of application in transportation problems. These areas of application are identified by three main classes: planning and design problems, traffic flow estimation problems and short term travel time prediction problems. The most important improvement is certainly in their better estimation power with respect to the available analytic formulas for planning purposes. Moreover, the knowledge of the variability of such queues (and consequently of the overflow delays) can be very important if the utility of a route alternative depends on the uncertainty of the travel time drivers may experience in their next trip. Finally contributions can be found in the short term travel time predictions, since these predictions strongly depend on how general is the model in terms of correctness of its outcomes and of strictness of its simplifying assumptions. In this sense the newly developed models can help at refining the errors coming from the data collection method and to validate or complete such datasets.

The author envisages that the proposed probabilistic modeling approach opens new perspectives in the development of dynamic and stochastic travel time estimation problems and it is a suitable methodology for design, planning and reliability analyses.

Samenvatting

Congestie neemt in stedelijke netwerken steeds serieuzere vormen aan en pieken in de verkeersvraag zijn steeds van langere duur. Een manier om dit probleem te verkleinen is door de infrastructuur aan te passen, bijv. door het toevoegen van nieuwe stroken of het aanleggen van nieuwe wegen. Daarnaast kunnen verkeersmanagers de netwerk condities ook verbeteren door het beschikbare netwerk efficiënter te gebruiken. Enkele van deze alternatieve management strategieën vallen onder de noemer Dynamisch Verkeers Management (DVM). Het doel van DVM is om de verkeersveiligheid en het gebruik van de infrastructuur te verbeteren. Dit wordt gerealiseerd door de dynamische aanpassing van de beschikbare capaciteit aan de vraag (bijv. door de toepassing van adaptieve verkeerssignalen) of vice versa (bijv. dynamische tolheffing). Om de inzet van deze DVM maatregelen te optimaliseren zijn zowel verkeersmodellen nodig om de impact van de maatnamen te voorspellen als ook om de maatregels te vinden, die de prestaties van het verkeerssysteem optimaliseren. Onder andere vormt de verhoging van de betrouwbaarheid een belangrijke doelstelling, die gerelateerd is aan veiligheid en capaciteit. Derhalve wordt kennis omtrent de variabiliteit van prestatiematen, zoals reistijden, vertragingen, etc. ook belangrijker.

Kennis omtrent de variabiliteit van reistijden is van fundamenteel belang in de evaluatie van DVM strategieën, maar speelt daarnaast ook een belangrijke rol in de keuzes van reizigers. Zo nemen reizigers reistijdvariabiliteit soms zelfs meer in acht dan gemiddelde reistijden. Vooral als er tijdsbeperkingen gelden (bijv. belangrijke vergaderingen) zijn weggebruikers zeer bezorgd over het risico van onverwacht lange reistijden en zijn ze daarom bereid een route te kiezen, die gemiddeld langer in beslag neemt, maar die voorspelbaarder is en daarmee de kans om te laat te komen vermindert. Zowel het belang dat tegenwoordig door verkeersanalisten en de overheid wordt gegeven aan de betrouwbaarheid van reistijden en het verminderen van de onzekerheid als de gevolgen van reistijdvariabiliteit op weggebruikers, vormen een belangrijke reden voor onderzoek van dit aspect van reistijden naast het onderzoek van gemiddelde reistijden.

Reistijden worden in stedelijke netwerken voor een groot deel bepaald door de vertraging, die bestuurders oplopen bij (geregelde) kruispunten; vandaar dat dit proefschrift zich voornamelijk toespitst op vertragingen opgelopen door verkeerslichten en hun verdelingen, de methodes die gebruikt zijn om de modellen te ontwikkelen

kunnen echter ook worden toegepast op verschillende andere operationele en gedragsmatige onderwerpen.

De wetenschappelijke wereld is het er over eens dat er nog steeds grote verbeteringen nodig zijn in het modelleren van vertragingen, bovenal door het variabele aspect van het vormen en verdwijnen van wachtrijen zowel binnen een cyclus als van cyclus tot cyclus. Het proces van de wachtrijvorming op met lichten geregelde kruispunten wordt al een halve eeuw bestudeerd. Desalniettemin blijft het een zwak punt in de verkeersstroomtheorie van stedelijke netwerken. De lange lijst van in het verleden ontwikkelde heuristieken getuigt van het feit dat een veelomvattende theorie, die kan uitleggen hoe dergelijke systemen werken, nog steeds ontbreekt.

Gegeven de probabilistische aard van verkeer, bestaan grote onzekerheden rond de schatting en voorspelling van de werking van verkeersregelingen. Eenvoudig gezegd kunnen alleen probabilistische modellen deze onzekerheid beschrijven en de dynamiek van wachtrijen en vertragingen representeren zonder de lange termijn effecten van de variabiliteit van verkeer te verwaarlozen.

Probabilistische modellen gebruiken juiste macroscopische relaties tussen de toestand, de regelvariabelen en de resulterende prestatiematen en nemen aan dat deze variabelen statistisch verdeeld zijn volgens een bekende kansverdeling. Dientengevolge worden ook de prestatiematen op een probabilistische wijze berekend. Een probabilistische benadering biedt de mogelijkheid om de statistische eigenschappen van verkeer te analyseren en om schattingen te geven van zowel de verwachte condities als de variabiliteit van verkeer door middel van de berekening van de standaard deviatie of 10-90% betrouwbaarheidswaarden.

Uitgaande van een eerder ontwikkeld model voor overbelastingswachtrijen van vaste-tijdenregelingen gebaseerd op de renewal-theorie van Markov ketenprocessen, is een probabilistisch model ontwikkeld dat het proces van wachtrijvorming ook binnen elke cyclus beschrijft. Dit model maakt het mogelijk om de lengte van de wachtrij te bepalen op elk tijdstip binnen een cyclus; het geeft uitbreiding aan het eerder ontwikkelde Markov model dat de verdeling van de lengte van de wachtrij berekent op basis van discrete tijdsstappen en het rechtvaardigt het dynamische en stochastische karakter van overbelastingswachtrijen. Het Markov ketenproces dat in hoofdstuk 4 is gepresenteerd, representeert een zeer krachtige techniek voor de modellering van dynamische en stochastische processen, omdat het variabelen op een probabilistisch niveau benadert en het verkeer simuleert met behulp van exacte uitdrukkingen gebaseerd op massabalans vergelijkingen. Deze eigenschap van de berekeningen maakt deze modellen voor planningsdoeleinden meer geschikt dan microscopische modellen, omdat ze de variabiliteit van verkeer simuleren en oorzaak-gevolg relaties tussen toestand, regelvariabelen en prestatiematen analyseren binnen aanvaardbare rekentijden.

Het nieuw ontwikkelde model is afgeleid zonder enige veronderstellingen omtrent de statistische eigenschappen van het aankomstproces. Het is zowel geschikt voor Poisson, binomiale en normaal verdeelde aankomsten. Belangrijker dan de mogelijkheid om de willekeurige vertraging te berekenen met een meer generiek model dan de reeds bestaande is het inzicht dat de afleiding geeft in het proces dat de willekeur in de vertraging veroorzaakt. De veronderstelling, die enkele auteurs maken dat de wachtrij weergegeven moet worden als een stapfunctie blijkt overtollig. Het stapsgewijze karakter van de vertraging wordt getransformeerd in een glad karakter van de verwachte wachtrijlengte, lineair toenemend in de roodfase en het eerste gedeelte van de groenfase. De verwachte waarde van de wachtrij in de groenfase toont een niet-lineair karakter op het moment dat de staart van de kansverdeling dicht bij nul komt. Dit fenomeen veroorzaakt overbelastingsvertraging.

Deze methode heeft de fout gekwantificeerd, die standaard analytische procedures als de HCM 2000 maken in de berekening van voertuigvertragingen. Dit geldt vooral wanneer het model afnemende wachtrijen moet berekenen na een oververzadigde periode, wat frequent voorkomt in piekperioden. Alle momenteel beschikbare methodes onderschatten vertragingen van voertuigen, of vervlakken vertragingen in een evaluatieperiode, in plaats van meer nauwkeurig vertragingen te berekenen voor elke cyclus. Een gecombineerde analyse van de gemiddelde tijd in een wachtrij over de tijd en de standaard deviatie hiervan toont aan dat er een sterke onderlinge afhankelijkheid bestaat tussen deze twee karakteristieken, vooral in verzadigde verkeerscondities. Dergelijke verkeerscondities doen een vertraging ontstaan die zich voortbeweegt door de tijd en die extra wachttijden doet ontstaan voor voertuigen, die de kruising naderen zelfs enige cycli nadat congestie is ontstaan.

Een exponentieel gedrag wordt gevonden voor de situatie waarin wachtrijen herstellen van lange initiële lengtes en het verkeerslicht bijna aan zijn capaciteit zit. Deze bevinding heeft geleid tot een nieuwe formulering voor de dynamiek van overbelastingswachtrijen, die de Van Zuylen-Viti formule genoemd wordt. Deze formulering combineert het deterministische lineaire karakter met een glad aansluitend asymptotisch gedrag in de buurt van de evenwichtswaarde. Daarnaast wordt een uitdrukking voor de standaard deviatie van de overbelastingswachtrij voorgesteld. De voorgestelde modellen kunnen wachtrijen en vertragingen berekenen onder zowel de veronderstelling van stationaire als niet-stationaire vraag. Deze eigenschap maakt dat ze bijvoorbeeld geschikt zijn voor “Dynamic Loading” processen, maar ook voor modelgebaseerde voorspellingen van verwachte wachttijden en voor plannings- en ontwerpproblemen.

De Markov en de Van Zuylen-Viti formules zijn vergeleken met de resultaten van het microsimulatie programma VISSIM, hierbij bleek een goede overeenkomst te bestaan. De consistentie tussen de drie benaderingen in verschillende verkeerscondities valideert de twee minder gedetailleerde methodes. Dit vormt ook een belangrijke bijdrage voor verkeersmanagers en mensen uit de praktijk, omdat het aantoont dat de dynamiek van overbelastingswachtrijen goed benaderd wordt met alle drie de modellen, ondanks de verschillen in gedetailleerdheid.

De probabilistische benadering biedt ook de mogelijkheid om wachtrijen te modelleren onder ruimere veronderstellingen dan de aannames van eenvoudige vaste-tijdenregeling en enkel-strooks geïsoleerde kruisingen. Terwijl er weinig verschil is tussen het formuleren van een Markov model voor geïsoleerde kruispunten en kruispunten in een netwerk (namelijk de aankomst verdeling), blijkt het een stuk lastiger te zijn om de samenhang tussen de strookkeuze van automobilisten en de lengte van de wachtrij te modelleren. Om rekening te houden met deze samenhang is het Markov model gecombineerd met een probabilistisch strookwisselmodel. Zodoende is aangetoond dat de verdeling van de aankomsten een dynamisch karakter heeft, in lijn met het dynamische karakter van de lengte van de overbelastingswachtrij. Daarnaast, maakt het Markov model voor meerstrooks kruispunten het mogelijk om rekening te houden met terugslag effecten, wat nuttige informatie biedt voor een correcte schatting van vertragingen en voor het ontwerp van aparte stroken voor afslaand verkeer. Tenslotte is de aanname van vaste-tijdenregeling versoepeld door de formulering van een probabilistisch model voor voertuigafhankelijke regeling. Deze benadering biedt de mogelijkheid om de kans op verlenging van de groentijd te berekenen, die afhangt van de variabiliteit van de aankomsten en hun volgtijdverdeling over de tijd.

De nieuw ontwikkelde modellen, die in deze dissertatie gepresenteerd worden, vormen belangrijke bijdragen op verschillende toepassingsgebieden in transport gerelateerde problemen. Deze toepassingsgebieden zijn in drie hoofdstromen te verdelen: plannings- en ontwerpproblemen, verkeersstroomschattingsproblemen en korte termijn reistijdvoorspellingen. De belangrijkste verbetering ligt zonder meer in de kracht van betere schattingen voor planningsdoeleinden ten opzichte van de beschikbare analytische formules. Verder kan de kennis betreffende de variabiliteit van dergelijke wachtrijen (en daarmee ook de variabiliteit van de vertragingen ten gevolge van overbelasting) zeer belangrijk zijn wanneer de waarde die toegekend wordt aan een route afhangt van de onzekerheid over de reistijd die bestuurders kunnen ervaren in hun volgende trip. Tenslotte is er een bijdrage in de korte-termijnvoorspelling van reistijden, omdat deze voorspellingen sterk afhankelijk zijn van de correctheid van de uitkomsten die het model geeft en de striktheid van de gemaakte vereenvoudigende veronderstellingen. In dit licht kunnen de nieuw ontwikkelde modellen hulp bieden in het opschonen van meetdata en het valideren en completeren van dergelijke datasets.

De auteur voorziet dat de voorgestelde probabilistische modelleringsbenadering nieuwe perspectieven opent voor dynamische en stochastische reistijdvoorspellingsproblemen en dat het een geschikte methode is voor ontwerps-, plannings- en betrouwbaarheids analyses.

TRAIL Thesis Series

A series of The Netherlands TRAIL Research School for theses on transport, infrastructure and logistics.

Nat, C.G.J.M., van der, *A Knowledge-based Concept Exploration Model for Submarine Design*, T99/1, March 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Westrenen, F.C., van, *The Maritime Pilot at Work*: E*valuation and Use of a Time-to-boundary Model of Mental Workload in Human-machine Systems*, T99/2, May 1999, TRAIL Thesis Series, Eburon, The Netherlands

Veenstra, A.W., *Quantitative Analysis of Shipping Markets,* T99/3, April 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Minderhoud, M.M., *Supported Driving: Impacts on Motorway Traffic Flow,* T99/4, July 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Hoogendoorn, S.P., *Multiclass Continuum Modelling of Multilane Traffic Flow*, T99/5, September 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Hoedemaeker, M., *Driving with Intelligent Vehicles: Driving Behaviour with Adaptive Cruise Control and the Acceptance by Individual Drivers*, T99/6, November 1999, TRAIL Thesis Series, Delft University Press, The Netherlands

Marchau, V.A.W.J., *Technology Assessment of Automated Vehicle Guidance - Prospects for Automated Driving Implementation*, T2000/1, January 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Subiono, *On Classes of Min-max-plus Systems and their Applications*, T2000/2, June 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Meer, J.R., van, *Operational Control of Internal Transport*, T2000/5, September 2000, TRAIL Thesis Series, Delft University Press, The Netherlands

Bliemer, M.C.J., *Analytical Dynamic Traffic Assignment with Interacting User-Classes: Theoretical Advances and Applications using a Variational Inequality Approach,* T2001/1, January 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Muilerman, G.J., *Time-based logistics: An analysis of the relevance, causes and impacts,* T2001/2, April 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Roodbergen, K.J., *Layout and Routing Methods for Warehouses*, T2001/3, May 2001, TRAIL Thesis Series, The Netherlands

Willems, J.K.C.A.S., *Bundeling van infrastructuur, theoretische en praktische waarde van een ruimtelijk inrichtingsconcept*, T2001/4, June 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Binsbergen, A.J., van, J.G.S.N. Visser, *Innovation Steps towards Efficient Goods Distribution Systems for Urban Areas,* T2001/5, May 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Rosmuller, N., *Safety analysis of Transport Corridors,* T2001/6, June 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Schaafsma, A., *Dynamisch Railverkeersmanagement, besturingsconcept voor railverkeer op basis van het Lagenmodel Verkeer en Vervoer,* T2001/7, October 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Bockstael-Blok, W., *Chains and Networks in Multimodal Passenger Transport. Exploring a design approach*, T2001/8, December 2001, TRAIL Thesis Series, Delft University Press, The Netherlands

Wolters, M.J.J., *The Business of Modularity and the Modularity of Business*, T2002/1, February 2002, TRAIL Thesis Series, The Netherlands

Vis, F.A., *Planning and Control Concepts for Material Handling Systems*, T2002/2, May 2002, TRAIL Thesis Series, The Netherlands

Koppius, O.R., *Information Architecture and Electronic Market Performance,* T2002/3, May 2002, TRAIL Thesis Series, The Netherlands

Veeneman, W.W., *Mind the Gap; Bridging Theories and Practice for the Organisation of Metropolitan Public Transport*, T2002/4, June 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Nes, R. van, *Design of multimodal transport networks, a hierarchical approach,* T2002/5, September 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Pol, P.M.J., *A Renaissance of Stations, Railways and Cities, Economic Effects, Development Strategies and Organisational Issues of European High-Speed-Train Stations*, T2002/6, October 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Runhaar, H., *Freight transport: at any price? Effects of transport costs on book and newspaper supply chains in the Netherlands*, T2002/7, December 2002, TRAIL Thesis Series, Delft University Press, The Netherlands

Spek, S.C., van der, *Connectors. The Way beyond Transferring*, T2003/1, February 2003, TRAIL Thesis Series, Delft University Press, The Netherlands

Lindeijer, D.G., *Controlling Automated Traffic Agents*, T2003/2, February 2003, TRAIL Thesis Series, Eburon, The Netherlands

Riet, O.A.W.T., van de, *Policy Analysis in Multi-Actor Policy Settings. Navigating Between Negotiated Nonsense and Useless Knowledge*, T2003/3, March 2003, TRAIL Thesis Series, Eburon, The Netherlands

Reeven, P.A., van, *Competition in Scheduled Transport*, T2003/4, April 2003, TRAIL Thesis Series, Eburon, The Netherlands

Peeters, L.W.P., *Cyclic Railway Timetable Optimization*, T2003/5, June 2003, TRAIL Thesis Series, The Netherlands

Soto Y Koelemeijer, G., *On the behaviour of classes of min-max-plus systems*, T2003/6, September 2003, TRAIL Thesis Series, The Netherlands

Lindveld, Ch..D.R., *Dynamic O-D matrix estimation: a behavioural approach*, T2003/7, September 2003, TRAIL Thesis Series, Eburon, The Netherlands

Weerdt, de M.M., *Plan Merging in Multi-Agent Systems*, T2003/8, December 2003, TRAIL Thesis Series, The Netherlands

Langen, de P.W, *The Performance of Seaport Clusters*, T2004/1, January 2004, TRAIL Thesis Series, The Netherlands

Hegyi, A., *Model Predictive Control for Integrating Traffic Control Measures*, T2004/2, February 2004, TRAIL Thesis Series, The Netherlands

Lint, van, J.W.C., *Reliable Travel Time Prediction for Freeways*, T2004/3, June 2004, TRAIL Thesis Series, The Netherlands

Tabibi, M., *Design and Control of Automated Truck Traffic at Motorway Ramps*, T2004/4, July 2004, TRAIL Thesis Series, The Netherlands

Verduijn, T. M., *Dynamism in Supply Networks: Actor switching in a turbulent business environment*, T2004/5, September 2004, TRAIL Thesis Series, The Netherlands

Daamen, W., *Modelling Passenger Flows in Public Transport Facilities*, T2004/6, September 2004, TRAIL Thesis Series, The Netherlands

Zoeteman, A., *Railway Design and Maintenance from a Life-Cycle Cost Perspective: A Decision-Support Approach*, T2004/7, November 2004, TRAIL Thesis Series, The Netherlands

Bos, D.M., *Changing Seats: A Behavioural Analysis of P&R Use,* T2004/8, November 2004, TRAIL Thesis Series, The Netherlands

Versteegt, C., *Holonic Control For Large Scale Automated Logistic Systems*, T2004/9, December 2004, TRAIL Thesis Series, The Netherlands

Wees, K.A.P.C. van, *Intelligente voertuigen, veiligheidsregulering en aansprakelijkheid. Een onderzoek naar juridische aspecten van Advanced Driver Assistance Systems in het wegverkeer,* T2004/10, December 2004, TRAIL Thesis Series, The Netherlands

Tampère, C.M.J., *Human-Kinetic Multiclass Traffic Flow Theory and Modelling: With Application to Advanced Driver Assistance Systems in Congestion*, T2004/11, December 2004, TRAIL Thesis Series, The Netherlands

Rooij, R.M., *The Mobile City. The planning and design of the Network City from a mobility point of view,* T2005/1, February 2005, TRAIL Thesis Series, The Netherlands

Le-Anh, T., *Intelligent Control of Vehicle-Based Internal Transport Systems*, T2005/2, April 2005, TRAIL Thesis Series, The Netherlands

Zuidgeest, M.H.P., *Sustainable Urban Transport Development: a Dynamic Optimization Approach*, T2005/3, April 2005, TRAIL Thesis Series, The Netherlands

Hoogendoorn-Lanser, S., *Modelling Travel Behaviour in Multimodal Networks*, T2005/4, May 2005, TRAIL Thesis Series, The Netherlands

Dekker, S., *Port Investment – Towards an integrated planning of port capacity*, T2005/5, June 2005, TRAIL Thesis Series, The Netherlands

Koolstra, K., *Transport Infrastructure Slot Allocation*, T2005/6, June 2005, TRAIL Thesis Series, The Netherlands

Vromans, M., *Reliability of Railway Systems*, T2005/7, July 2005, TRAIL Thesis Series, The Netherlands

Oosten, W., *Ruimte voor een democratische rechtsstaat. Geschakelde sturing bij ruimtelijke investeringen*, T2005/8, September 2005, TRAIL Thesis Series, Sociotext, The Netherlands

Le-Duc, T., *Design and control of efficient order picking*, T2005/9, September 2005, TRAIL Thesis Series, The Netherlands

Goverde, R., *Punctuality of Railway Operations and Timetable Stability Analysis*, T2005/10, October 2005, TRAIL Thesis Series, The Netherlands

Kager, R.M., *Design and implementation of a method for the synthesis of travel diary data*, T2005/11, October 2005, TRAIL Thesis Series, The Netherlands

Boer, C., *Distributed Simulation in Industry*, T2005/12, October 2005, TRAIL Thesis Series, The Netherlands

Pielage, B.A., *Conceptual Design of Automated Freight Transport Systems*, T2005/14, November 2005, TRAIL Thesis Series, The Netherlands

Groothedde, B., *Collaborative Logistics and Transportation Networks, a modeling approach to network design*, T2005/15, November 2005, TRAIL Thesis Series, The Netherlands

Valk, J.M., *Coordination among Autonomous Planners*, T2005/16, December 2005, TRAIL Thesis Series, The Netherlands

Krogt, R.P.J. van der, *Plan Repair in Single-Agent and Multi-Agent Systems,* T2005/17, December 2005, TRAIL Thesis Series, The Netherlands

Bontekoning, Y.M., *Hub exchange operations in intermodal hub-and-spoke networks. A performance comparison of four types of rail-rail exchange facilities*,T2006/1, February 2006, TRAIL Thesis Series, The Netherlands

Lentink, R., *Algorithmic Decision Support for Shunt Planning*, T2006/2, February 2006, TRAIL Thesis Series, The Netherlands

Ngoduy, D., *Macroscopic Discontinuity Modeling for Multiclass Multilane Traffic Flow Operations,* T2006/3, April 2006, TRAIL Thesis Series, The Netherlands

Vanderschuren, M.J.W.A., *Intelligent Transport Systems for South Africa. Impact assessment through microscopic simulation in the South African context,* T2006/4, August 2006, TRAIL Thesis Series, The Netherlands

Ongkittikul, S., *Innovation and Regulatory Reform in Public Transport,* T2006/5, September 2006, TRAIL Thesis Series, The Netherlands

Yuan, J., *Stochastic Modelling of Train Delays and Delay Propagation in Stations*, T2006/6, October 2006, TRAIL Thesis Series, The Netherlands

Viti, F., *The Dynamics and the Uncertainty of Delays at Signals*, T2006/7, November 2006, TRAIL Thesis Series, The Netherlands