

SUSTAINABLE URBAN TRANSPORT DEVELOPMENT

A DYNAMIC OPTIMISATION APPROACH

Dissertation committee:

prof. dr. ir. H. J. Grootenboer	University of Twente, chairman/secretary
prof. dr. ir. M. F. A. M. van Maarseveen	University of Twente, promotor
prof. dr. E. O. Akinyemi	University of Ilorin, Nigeria, promotor
prof. dr. ir. B. van Arem	University of Twente
prof. dr. ir. E. C. van Berkum	University of Twente
prof. dr. G. P. M. R. Dewulf	University of Twente
prof. dr.-ing. H. Knoflacher	Vienna University of Technology, Austria
dr. S. P. Shepherd	University of Leeds, United Kingdom

Trail Thesis Series T2005/3, The Netherlands TRAIL Research School

This thesis is the result of a Ph.D. study carried out between 1999 and 2005 at the University of Twente, faculty of Engineering Technology, department of Civil Engineering, Centre for Transport Studies.

Cover picture: Detail of rickshaw and taxis in a Kolkata street, India
Copyright © by Bennett Dean/Eye Ubiquitous/Corbis/TCS

Typeset in L^AT_EX

Copyright © 2005 by M.H.P. Zuidgeest, Enschede, The Netherlands

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the written permission of the author.

Printed by Febodruk BV, Enschede, The Netherlands.

ISBN 90-365-2174-2

SUSTAINABLE URBAN TRANSPORT DEVELOPMENT

A DYNAMIC OPTIMISATION APPROACH

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de Rector Magnificus,
prof. dr. W.H.M. Zijm,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op donderdag 28 april 2005 om 16.45 uur

door

MARCUS HENRICUS PETRUS ZUIDGEEST

geboren op 23 januari 1974
te Den Helder

Dit proefschrift is goedgekeurd door de promotoren:

prof. dr. ir. M. F. A. M. van Maarseveen

prof. dr. E. O. Akinyemi

“It is twenty-five kilometres from Addis Ababa to the Sabeta waterfall. Driving a car in Ethiopia is a kind of unending process of compromise: everyone knows that the road is narrow, old, crammed with people and vehicles, but they also know that they must somehow find a spot for themselves on it, and not only find a spot, but actually move, advance forward, make their way toward their destination. Every few moments, each driver, cattle herder, or pedestrian is confronted by an obstacle, a conundrum, a problem that needs solving: how to pass without colliding with the car approaching from the opposite direction; how to hurry along one’s cows, sheep, and camels without trampling the children and crawling beggars; how to cross without getting run over by a truck, being impaled on the horns of a bull, knocking over that woman carrying a twenty-kilogram weight on her head. And yet no one shouts at anyone else, no one falls into a fury, no one curses or threatens-patiently and silently, they all perform their slalom, execute their pirouettes, dodge and evade, maneuver and hedge, turn here, converge there, and, most important, move forward.”

Ryszard Kapuściński
The Shadow of the Sun: my African life.

Preface

In my opinion, Ryszard Kapuściński's observations on driving a car in Ethiopia (page *v*) very well symbolise the process of a Ph.D. research. As a matter of fact, the same observations also perfectly stand for my personal motivation to start a research in modelling sustainable transport, back in 1999, in the first place.

When I was walking in a tropical city for the first time in my life, that was in Guatemala-City in 1995, its traffic fascinated me very much. Busyness and noise. Chaotic traffic, but still structured and often very efficient. Understanding, describing and modelling this apparent chaos, that would be the topic of my research! In the end, everything turned out to be different, and this research got a more general character. Still, I hope this work motivates many people to continue research in the interesting and important field of transport sustainability.

In a traffic flow you are never alone, hence I would like to mention and thank some people who, sometimes literally, travelled along with me.

First and foremost, I am indebted to my supervisors, Prof. Dr. Martin van Maarseveen and Prof. Dr. Edward Akinyemi, who provided the necessary (intellectual) infrastructure and means to do research. Eddie, your qualities and qualifications are often underestimated, I am proud to say I did my Ph.D. with you. Thank you for all your valuable lessons (in life). Martin, in the famous Hôtel le Bénin in Lomé some dodgy characters thought you were my father. Our conversations were indeed often very personal, almost family like, where you showed great insight in human nature. Thank you for the confidence in me.

In the early stages of the dynamic modelling, I had the privilege and pleasure to work with Dr. Clifford Wymer, former econometrist with the International Monetary Fund. Clifford, your enthusiasm for continuous-time macro-economic modelling and your never-ending patience with me are least-to-say unique. Thank you very much. Also the hospitality I received in London while staying with you and your wife Jill are unforgettable. It is a pity I had to move away from using your software in the end.

The discussions I had with Dr. Kieran Donaghy from the University of Illinois at Urbana-Champaign, USA, and Dr. Laurie Schintler from George Mason University, USA, during a workshop in Bonn and by e-mail were very useful for better understanding of their interesting model. Thank you very much.

My colleagues in the UT department of mathematical systems and control theory helped me a lot, especially with respect to mathematical notations. In particular, I wish to acknowledge Dr. Bram van den Broek.

I would also like to thank the members of the dissertation committee for reading the thesis and giving me their criticisms and suggestions.

During my research visit to the National Center for Transportation Studies, University of the Philippines Diliman in Manila, I received great hospitality and help, in particular by Dr. Noriel Tiglao, Dr. Jose Regidor, Mr. Ed Kamid, Ms. Alorna Abao and Dr. Ricardo Sigua. Furthermore, in the Asian Development Bank I was kindly received by Mr. Herbert Fabian and Mr. Cornie Hui-zenga. In addition, I would like to thank my colleagues and students in ITC, in particular Mr. Rizal Cruz, Mr. Javier Pacheco, Mr. Sherif Amer, Mr. Johan de Meijere as well as Dr. Luc Boerboom, who helped me a lot in organising the field visits to Malaysia and The Philippines.

The contributions of some of my former M.Sc. students who tackled interesting problems in the field of sustainable transport that were of direct or indirect importance to this thesis are also greatly acknowledged. In particular I would like to mention the researches of Mr. Thijs Oude Moleman on sustainability indicators, Mr. Manus Barten on optimisation in transport planning (partly reported in chapter 4), Mr. Alex van Gent on transport network aggregation, Ms. Annette van Nes on environmental capacity, Mr. Roland Kager on transport modelling in developing countries and Mr. Martijn Bierman on the existence and measurement of elastic demand in Metro Manila.

Also thanks to (former) colleagues and friends from within the UT faculty of Engineering Technology, in particular to Giovanni (for keeping me company during weekend and evening shifts), Peter & Mako (for the many dinners and long evenings in café Bolwerk), Bas & Roland (for putting things in perspective with their Centre of Irrelevance), Frans & Kasper (for stepping in with Verkeer as well as being great colleagues), Bart (2×), Martin, Eric, Henny, Wendy, Thijs, Cornelia, Mascha, Cees, Michiel, Jean-Luc, Jebbe, Godfried, Attila, Pieter, Jörg, Caroline, Jan-Willem, Wilbert, Astrid, Marjolein, Suzanne, Leo, Marc (for reviewing chapter 4), Maureen, Dorette, Harry, Henriëtte, Graziana, Hans, Sam, Matthijs, Anne, René, Judith, Anne-Marie, Lynn, Rien, Denie, and Martijn (2×), who make that I still enjoy working here.

My former colleagues and students in UNESCO-IHE-Delft have introduced me to so many interesting topics as well as cultures. In particular, many thanks go to Mr. Jan Herman Koster, who kept on believing in me, even after I decided

to leave for Twente.

My (former) colleagues in Keypoint Consultancy, in particular Mr. Leo de Jong, Dr. Marc Witbreuk and Mr. Lars Mosch, who made that once in a while I could focus on the practical side of traffic and transport by reviewing their consultancy reports. Thanks for the trust in me!

On a personal note, no one can get by without help of friends and family. Too many persons to mention individually, I wish to say: 'groep Utrecht', 'campusmongolen', dear family(-in-law) and study-friends, Thanks! Bart and Peter, thank you very much for being paranympths during my defence. Your friendship and help throughout the last five years are very much appreciated.

Finally, I would like to thank Cari, who, even when I seemed to walk around in a fog at times, stayed confident and optimistic. Cari, without your love and care this work would probably never have finished.

mz

Contents

Preface	vii
I Sustainable urban transport development	1
1 Introduction	3
1.1 Transport realities	3
1.2 Transport planning	8
1.2.1 Supply and demand	9
1.2.2 Transport modelling	12
1.2.3 Transport modelling alternatives	22
1.3 Strategic transport planning	24
1.4 Problem statement and research questions	27
1.5 Limitations of this research	29
1.6 Scope and outline of this thesis	29
2 Sustainable development and transport	31
2.1 The Limits to Growth	31
2.2 Sustainability and sustainable development	35
2.2.1 Three conceptions of sustainable development	37
2.2.2 Three dimensions of sustainable development	39
2.2.3 Three levels of discourse	40
2.2.4 Sustainable urban development	41
2.3 Sustainable urban transport development	43
2.3.1 A framework for sustainable transport development	45
2.3.2 A different transport planning paradigm	49
3 Sustainable transport development requirements	53
3.1 Introduction	53
3.2 Movement needs and transport system performance	53
3.2.1 Level-of-Service	55
3.2.2 Mobility and person travel	55
3.2.3 Accessibility	57
3.2.4 Equity	61
3.3 Transport and development	63

3.3.1	Dynamics in travel demand	65
3.4	Resource consumption and capacities	70
3.4.1	Measures for controlling transition paths	74
II	Modelling sustainable transport development	77
4	Optimisation as a tool for sustainable transport policy design	79
4.1	Transport policy and planning	79
4.1.1	The transport policy process	79
4.1.2	Decision-making approaches	81
4.1.3	Complex transport policy objectives	82
4.2	Transport planning as a design problem	83
4.2.1	Existing optimisation models	86
4.2.2	Transport planning as an optimisation problem	87
4.3	Methods and models for modelling sustainable transport	92
4.3.1	Indicators for transport sustainability	93
4.3.2	Scenario techniques and backcasting	94
4.3.3	Optimising transport networks	95
5	A dynamic optimisation model	101
5.1	Introduction	101
5.2	Dynamic optimisation	102
5.2.1	Continuous-time modelling	103
5.2.2	Disequilibrium modelling	104
5.3	A dynamic transport optimisation model	108
5.3.1	Transport dynamics	109
5.3.2	Transport policy objectives	124
5.3.3	Measures and bounds	127
5.4	Pontryagin's Maximum Principle	129
5.4.1	Transversality conditions and path constraints	134
5.4.2	General computational aspects	136
5.4.3	Deriving the optimal control	138
III	Strategic modelling examples	145
6	Case studies	147
6.1	Introduction	147
6.2	Solving the optimal control in Matlab	147
6.3	Networks and cases	150
6.4	Case studies: congestion minimisation (C1)	152
6.4.1	An optimal U^c -control for case 1	153
6.4.2	An optimal U^c -control for case 2	158
6.4.3	An optimal U^t -control for case 2	163
6.4.4	An optimal U^c & U^m -control for case 2	167
6.4.5	An optimal U^c -control with a state constraint for case 2	177

6.4.6	An optimal U^c -control for case 3	184
6.5	Case studies: accessibility maximisation (C2)	190
6.5.1	An optimal U^c -control	191
6.5.2	An optimal U^c -control with an integral state constraint	195
6.6	Case study: person throughput maximisation (C3)	201
6.6.1	An optimal U^t -control with an emission state constraint	202
6.7	Case study: equity maximisation (C4)	207
6.7.1	An optimal U^t -control with an emission state constraint	207
6.8	Opportunities for a real-life application	211
7	Summary, conclusions and further research	215
7.1	Summary	215
7.2	Conclusions	217
7.2.1	The urge for sustainable urban transport development	218
7.2.2	Characterisation of the problem	219
7.2.3	Optimising transport policies	221
7.2.4	Model formulation and solution strategies	222
7.2.5	Sustainable urban transport development	228
7.3	Further research	230
7.3.1	Characterisation of the problem	230
7.3.2	Model formulation	231
7.3.3	Optimisation	232
7.3.4	Future application	233
8	Epilogue	235
A	Adverse effects from vehicle emissions	239
A.1	Air pollutants	239
A.2	Climate change	241
B	Models estimation	243
B.1	General least squares criterion	243
B.2	Maximum likelihood estimation	245
B.3	Estimating parameters in nonlinear differential systems	246
C	Further sufficiency conditions	249
C.1	Concave and convex functions	249
D	Partial derivatives	253
D.1	For optimal control problem A4/B3/C1	253
D.2	For optimal control problem A3/B1/C2	259
D.3	For optimal control problem A4/B2/C3	260
D.4	For optimal control problem A4/B2/C4	260
	Samenvatting (Dutch summary)	261
	Bibliography	264

Nomenclature	279
About the author	285
TRAIL Thesis Series	287

Part I

Sustainable urban transport development

Chapter 1

Introduction

1.1 Transport realities

Traffic and transport policies differ greatly from city to city, from country to country, as do the travel patterns of the people in these cities and countries. This is well explained by the differences in the social, political, economical as well as cultural context. Therefore, at first sight, an overcrowded dala-dala¹ negotiating the streets of Dar-es-Salaam, Tanzania, seems to have little in common with the Light Rail Transit system in Utrecht, The Netherlands, but both, despite the apparent disparities in their operations and technology, are fulfilling a basic demand for transport. Mobility and accessibility provided by the transport system have been playing a major role in shaping countries, influencing the location of social and economic activity, the form and size of cities, and the style and pace of life by facilitating trade, permitting access to people and resources, and enabling greater economies of scale, worldwide and throughout history. Furthermore, do they expand cultural and social connections, increase employment, and educational as well as healthcare opportunities.

Transport development aims at reducing time and energy, hence costs, spent on travel and transport, thereby improving people's access to resources, other people, freight, opportunities, markets and services they wish to reach. Unfortunately, one has to conclude that much of these transport development benefits are inequitably distributed spatially as well as socially. In many cities, especially those in developing countries, many people don't have access to adequate transport infrastructure and means of transport, because these are neither available nor affordable to them. People mostly walk, use bicycles or two-wheeled motorised vehicles, or depend on various forms of formal and informal public transport. Bicycles are limited in their range; two-wheeled motorised vehicles have a larger range, but are still expensive. Public transport is generally less expensive in terms of the out-of-pocket costs required to use it, but is often difficult to reach and provides relatively poor and inflexible service.

¹Informal means of public transport, 9-seater bus.

Table 1.1: Measures of transport infrastructure per capita [km mln.⁻¹ inh.⁻¹] (European Commission, 2000).

	Intercity rail	Urban rail	Roads	Motorways
European Union 15	415	18	9.330	125
Central and Eastern Europe	635	50+	7.880	24
United States	140 ^a /890	7	23.900	325
Japan	210	6	9.200	51
World	210	4	4.750	35

^aOnly 38.000 km in passenger service.

Inadequate infrastructure seriously impedes economic and social development. Extensive passenger rail networks exist only in Asia and Europe, and general roadway provision in developing countries falls far behind that in the developed world, see table 1.1. Lack of road capacity is therefore often a serious problem on urban roads. The basic connectivity of the road network may be deficient as well, with important population or economic centres poorly linked within the cities or to the rest of the country. In some cases, specific individual facilities such as bridges, footpaths or bicycle lanes are lacking too. Furthermore, the quality of road infrastructure is frequently not good, because of deficiencies in the original design and construction, but also inadequate control of trucks with excessive axle loads, inclement climatic conditions, or neglected maintenance (WBCSD, 2001).

There are, however, also some other distinct similarities between the developed and developing world; a repeating daily pattern: slow-moving queues of cars, trucks, buses, jeepney's and motorcycles, mixed with bicycles, rickshaws, handcarts, push-carts and pedestrians trying to move to centres of economic activity; at the same time trapping millions of people in an unsafe, noisy and polluted environment that is endangering flora and fauna, causing traffic accidents and serious traffic related health problems to people as well as taking away children's playgrounds. These observations accompanied by some impressive figures are given by John Whitelegg in *The Guardian* (Whitelegg, 2003): air pollution from traffic claims 400.000 lives each year, mostly in developing countries, and some 1.5 billion people are exposed every day to levels of pollution well in excess of World Health Organisation (WHO) recommended levels. Particulate pollution and levels of cancer-causing pollutants have already damaged the health of hundreds of millions of children. Table 1.2 shows some figures on air pollution levels in developing cities and some developed cities. The cities differ in the nature of the air pollution, as well as in the excess pollution produced, due to the specific characteristics of pollutant sources and the fuels used. However, in most cases transport is the main source of pollution (World Bank, 1996). In large city centres road traffic may account for as much as 90 to 95% of lead and carbon monoxide, 60 to 70% of nitrogen oxides and hydrocarbons, and a major share of particulate matter.

By 2030, it is predicted, 2.5 million people will be killed on the roads of develo-

Table 1.2: Large cities exceeding WHO pollution levels. Figures show concentration levels surpassing limits by a factor of up to 2 (< 2) or by more than 2 (> 2) (Vasconcellos, 2001).

City	Lead ^a	CO ₂ ^b	NO _x ^c	Ozone	SO ₂ ^d	PM ^e
Bangkok	< 2					> 2
Beijing				< 2	> 2	> 2
Bombay						> 2
Buenos Aires		n/a	n/a	n/a	n/a	< 2
Cairo	> 2	< 2				> 2
Kolkata		n/a		n/a		> 2
Delhi				n/a		> 2
Jakarta	< 2	< 2		< 2		> 2
Karachi	> 2	n/a	n/a	n/a		> 2
London		< 2				
Los Angeles		< 2	< 2	> 2		< 2
Manila	< 2	n/a	n/a	n/a		> 2
Mexico City	< 2	> 2	< 2	> 2	> 2	> 2
Moscow		< 2	> 2			< 2
New York		< 2		< 2		
Rio de Janeiro			n/a	n/a	< 2	< 2
São Paulo	< 2	< 2	< 2	> 2		< 2
Seoul					> 2	> 2
Shanghai	n/a	n/a	n/a	n/a	< 2	> 2
Tokyo				> 2		

^a90 - 100% from transport sources.

^bCarbon-dioxide, 80 - 100% from transport sources.

^cOxides of nitrogen, 60 - 70% from transport sources.

^dSulphur dioxide, 80 - 100% from transport sources.

^eSuspended particulate matter.

ping countries each year and 60 million people will be injured. Even now, 3.000 people are killed and 30.000 seriously injured on the world's roads every day. These deaths and injuries take place mainly to pedestrians, cyclists, bus users and children. The poor suffer disproportionately; they experience the worst air pollution and are deprived of education, health, water and sanitation programmes because the needs of the car soak up so much national income. Road transport absorbs massive public investments for building and maintenance. In cities as Kolkata, India and Nairobi, Kenya car ownership and use is growing at more than 20% a year, with little effort made to protect those not in cars. Advances in vehicle, engine and fuel technology are of little relevance in Asian and African cities, where the growth of car and lorry numbers is dramatic and where highly polluting diesel and two-stroke engine vehicles are the norm (Whitelegg, 2003). This rather sad picture is also pertinent in developed countries. In the European Union (EU) for example, passenger and freight transport have more than doubled between 1970 and 1997, with the strongest growth being in air and road transport, and are still growing. Here the car has increased its dominance at the expense of all other modes of transport, including the healthy green modes: the car increased its share of passenger transport from 65 to 74% between 1970 and 1997, and trucks now account for 45% of total freight transport compared with 30% in 1970 (see figure 1.1).

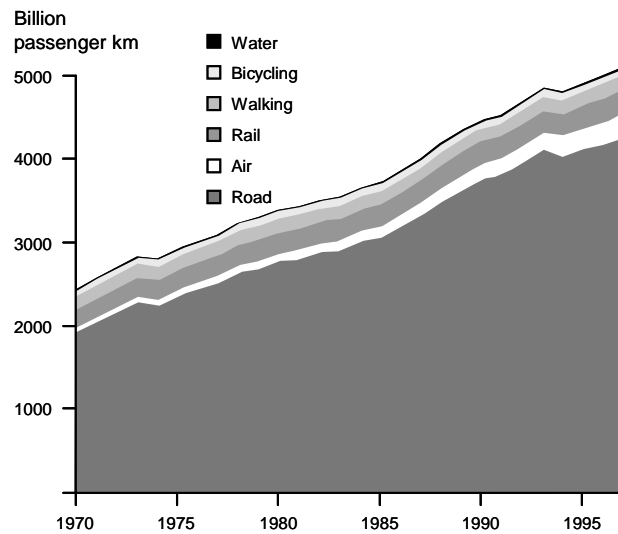


Figure 1.1: Increases in passenger travel in the European Union, 1970 - 1997 (Eurostat, 2001).

Between 1970 and 1997, passenger and freight transport in the EU increased by an annual average of 2.8 and 2.6% respectively, while the growth in Gross Domestic Product (GDP) over the same period was 2.5%. For road and air-passenger travel particularly, the boost in demand can be attributed to higher incomes, a fall in transport prices in real terms and changes in travel patterns, because of urban sprawl, decreasing household sizes, and changing work patterns and lifestyles. In turn, the demand and intensity of freight transport is closely linked to changes in the volume and structure of the economy and to infrastructure supply. Energy and carbon dioxide efficiency (i.e. energy use per passenger and per freight transport unit) has shown little or no improvement since the early 1970s (see figure 1.2)). The increasing use of heavier and more powerful vehicles, together with decreasing occupancy rates and load factors, has outweighed increases in vehicle energy efficiency due to technological advances. As a result, growing transport volumes led to about a 14% increase in energy consumption and a 12% increase in carbon dioxide emissions between 1990 and 1997; figures that are still pertinent today. On the positive side, emissions of non-methane volatile organic compounds (VOC) and nitrogen oxides have been falling since 1990 (also figure 1.2), mainly due to the introduction of catalytic converters in vehicle exhausts. However, the decrease has been slower than expected as increasing transport demand has partly offset engine improvements. Traffic noise is another key urban problem. It is estimated that over 30% of people in the EU are exposed to high road-traffic noise levels, about 10% of people to high rail noise levels, and possibly a similar proportion to aircraft noise (Eurostat, 2001). In addition, transport infrastructure takes land and may constitute a barrier against the movement of species (including

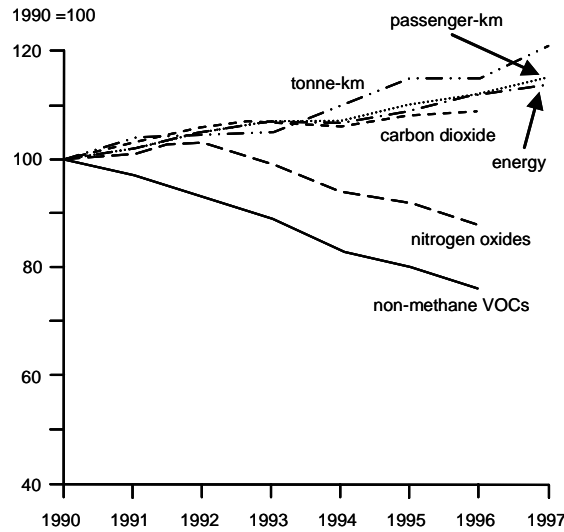


Figure 1.2: Transport eco-efficiency (air emissions) in the European Union (Eurostat, 2001).

man).

The mistaken notion, notably by transport professionals, that modernisation equals motorisation brought nothing more than congestion in cities, air pollution and traffic unsafety (Peters, 2002). Speeds during the morning peak periods in some congested urban areas around Europe have come down to a mere 17kmh^{-1} in the Central Business Districts (CBD), whereas in the built-up areas motorists still enjoy an average speed of 27kmh^{-1} . In less congested areas these figures are 23kmh^{-1} and 41kmh^{-1} respectively (ECMT, 1995); calculations indicate that the fatality rate per kilometre for walking and cycling in the United Kingdom (UK) is 15 to 12 times, respectively, the rate for car travel. Fatality rates for walking and cycling, compared with bus passengers, are even more extreme at 66 and 55 times respectively. The fact that walking and cycling are much less ‘safe’ forms of travel than car and bus, however, should not be taken as an encouragement to try to shift travellers from these ‘unsafe’ modes to ‘safer’ mechanised ones in order to reduce road casualties. A UK study, table 1.3 (next page) adopted from Hillman and Adams (1992), illustrates that the overall fatality rate associated with each kilometre of travel by different road users gives a different picture. The heaviest and sturdiest vehicles naturally run a very low risk of being killed in a road accident, but there is a very much greater risk that other road users will be killed, especially the more vulnerable ones. These figures suggest transport policy should be directed at those modes incurring the least threat to other road users, rather than the reverse.

The transport realities described here stand in large contrast to the paradigm

Table 1.3: Fatality rates in the UK per 1.0×10^8 vehkm to (Hillman and Adams, 1992):

Mode	Users themselves	Pedestrians	Other users	All users	Other as a % of all
Bicycle	4.9	0.1	0.1	5.1	4
Motorcycle	10.3	1.7	0.6	12.6	18
Car	0.7	0.4	0.4	1.5	53
Light freight	0.4	0.4	0.6	1.4	71
Bus	0.4	1.8	1.7	3.9	90
Heavy lorry	0.2	0.5	1.9	2.6	93

of a sustainable and developed transport system in which people have safe, quick and comfortable access to the activities they wish to perform while living in harmony with their natural environment on the short and long run. Even while some countries, notably the US administration, refuse, in contrast to many other countries, to ratify the Kyoto protocol - that aims at significantly reducing emission of Greenhouse gasses -, the Chinese government is banning out bicycles in their city centres - to make room for cars -, Mexico City, Beijing, Cairo, Jakarta, Los Angeles, São Paulo and Moscow (as seen in table 1.2) are competing to become number one polluted and congested megacities in the world, it is believed that sustainable urban transport development should be possible if a paradigm shift in thinking and acting is established, before sustainable transport development becomes a paranoia instead.

This thesis aims to demonstrate the implications of a transport planning paradigm shift and to develop a corresponding analytical framework involving change from a given state of the transport system to a system compatible with sustainable development. The design concept and its feasibility will also be discussed. The approach involved, combines elements from traditional reactive transport planning and precautionary elements of environmentalism to reach what Deike Peters calls a sustainable mobility consensus (Peters, 2002).

Next paragraph² discusses some basics on transport planning and transport modelling theory, necessary to understand the alternative modelling concept that will be introduced in paragraph 1.3.

1.2 Transport planning

Transport policies have changed over time, in response to for example increasing vehicle performance and ownership levels, increasing congestion as well as increasing awareness of environmental issues. Likewise, transport planning strategies and objectives have changed from fully satisfying demand for transport to a more objective oriented, but trial-and-error based, transport planning. Furthermore, as knowledge on travel behaviour of people in urban areas has improved, decision-makers, supported by transport modellers, were able

²Readers familiar with these theories may skip this paragraph up to section 1.2.3.

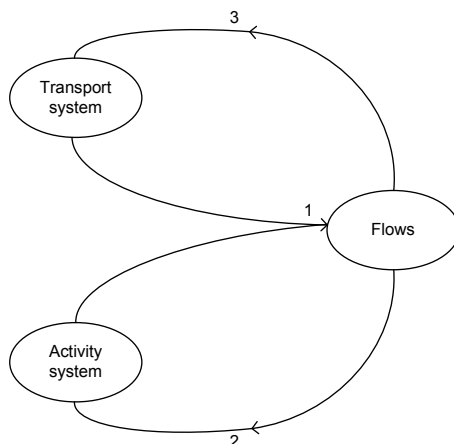


Figure 1.3: Basic relations of transport systems analysis (Manheim, 1979).

to change policy directions into management of road space. The basic issues involved in transport system analysis and transport planning are discussed in this paragraph.

1.2.1 Supply and demand

The transport system is closely linked to the social-economic system in an area. Back in the 1970s, Marvin Manheim posed the basic relations of transport system analysis; transport systems may shape societal and economic processes in the area. In return will these societal and economic processes indirectly affect the shape of the transport system (Manheim, 1979). In a systems approach three basic variables can be distinguished:

1. The transport system;
2. The activity system;
3. Traffic and transport flows.

Three basic relations between these variables are depicted in figure 1.3. First, the traffic and transport flows derive from the equilibration of transport system supply and the activity patterns that generate a travel demand. This is typically a short to medium term relationship. Second, the traffic and transport flows might on the longer-term change activity patterns (shifts in modal choice, trip frequency choice etceteras) and eventually land-use patterns. Third, traffic and transport flows might necessitate changes in the transport system itself, through actions of traffic managers and transport planners. Urban transport planning is concerned with this interaction, which is the equilibration of travel demand, i.e. derived demand from the activity system, and infrastructure supply, i.e. the characteristics of the transport system. It intends to steer the process of allocation of traffic generators and the provision of transport faci-

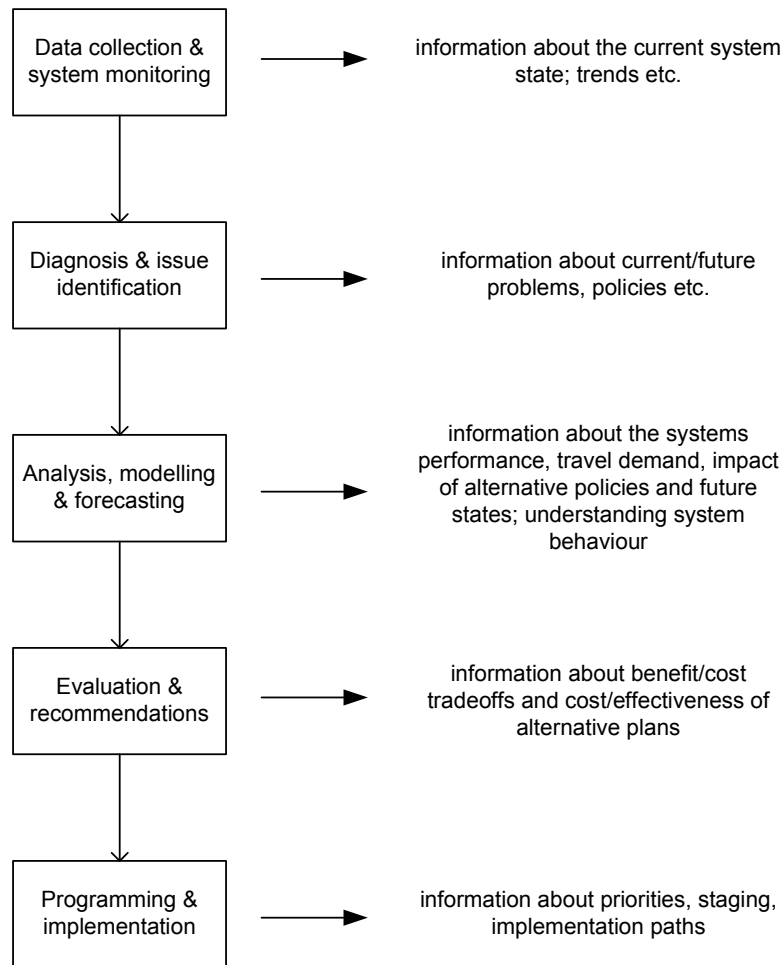


Figure 1.4: The transport planning process (Meyer and Miller, 2001).

ties over space and time (Tolley and Turton, 1995).

Urban transport planning therefore tries to control the equilibrium flow in Manheim's systems approach. The urban transport planning process is depicted in figure 1.4. Through the collection and monitoring of basic current and expected future traffic and transport related data as well as the analysis of these data, the transport problems at hand are derived in relation to (future) political, social and economic developments. In addition, quantitative techniques are applied to model, analyse and forecast alternative plans, and (future) scenarios. Several plans may then be evaluated on the basis of several indicators (cost-benefit analysis, environmental impact assessment etceteras). Consequently, one or a combination of plans might be implemented. Manheim's systems approach is typically applied in the analysis, modelling and forecasting

phase of this planning process. This is accordingly called the urban transport planning modelling system, or analytical transport planning model (TPM). As will become clear, this thesis focusses on the definition and operationalisation of such modelling techniques for the planning of sustainable urban transport systems.

The equilibration of infrastructure supply and travel demand is often depicted as a travel market equilibrium (as seen in figure 1.3). Both supply and demand are treated as functions of costs, in transport terms generalised costs, i.e. a (non)linear combination of weighted disutilities like travel cost (out-of-pocket cost and variable cost) and travel time converted to monetary units using the value-of-time concept, applying parameters α and β . For example, the generalised cost expressed as a linear weighted combination of travel time and travel cost reads:

$$c = \alpha \cdot \text{travel time} + \beta \cdot \text{travel cost.} \quad (1.1)$$

The notion that the demand for travel, T , is a function of cost, c , presents no difficulties. However, if the predicted travel demand were actually realised, the generalised cost might not stay constant. This is where the *infrastructure supply* model comes in. The classical approach defines the supply curve as giving the quantity T , which is produced, given a market price c . However, while certain aspects of the supply function do, of course, relate to the cost of providing services, the focus of supply relationships in transport has very often been on the non-monetary items, and on time in particular. In addition, the generalised cost is often more straightforward seen as the inverse relationship, whereby c is the unit (generalised) cost associated with meeting a demand T . The supply model thus reflects the response of the transport system to a given level of demand; that is the deterioration in speeds as traffic volumes rise, or increased parking problems as demand approaches capacity etceteras (Bates, 2001).

This equilibrium model is graphically depicted in figure 1.5 (next page). Considering both the demand curve and supply curve 1, the equilibrium point with actual travel, or revealed demand, can be found where both curves cross. This model can for example be used to see the effect of road expansion. Supply curve 2 indicates the changed situation as new capacity is added. At low demand volumes generalised costs c_1 do hardly change (non-congested area of the curve). However, at higher demands the generalised costs c_2 tend to rise slower than before due to the higher capacity. Hence, some extra demand is induced or generated, due to improved travel conditions. The latent demand is the demand that has not (yet) been revealed or become manifest, but could be revealed due to other transport system changes. As travel demand is a derived demand from activities, there will be a limit to the amount of latent demand available (a market saturation).

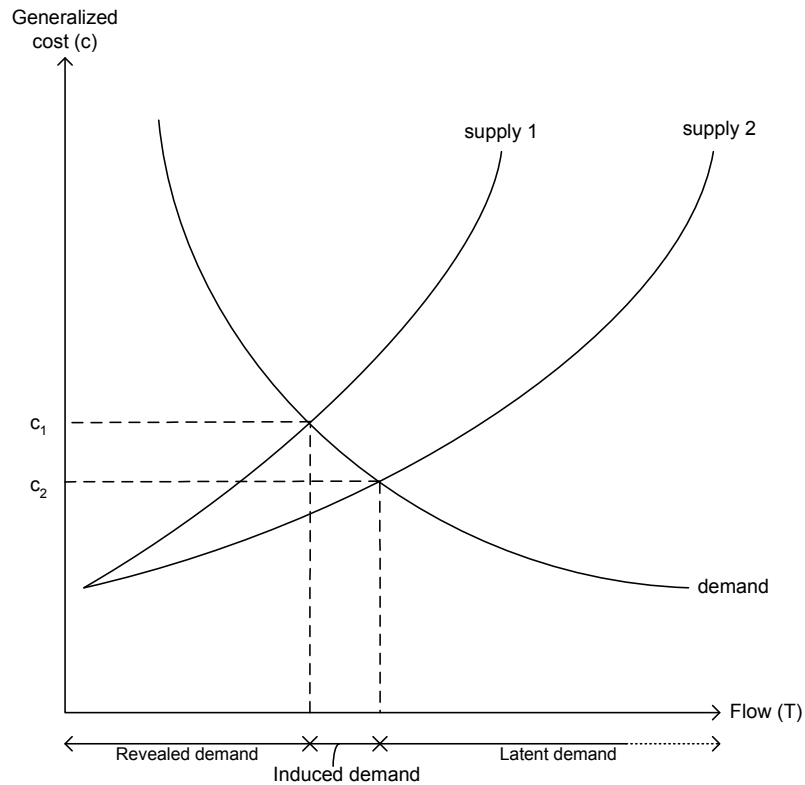


Figure 1.5: Demand-supply equilibria.

1.2.2 Transport modelling

The current approach in analytical transport planning or transport modelling, evolved from the 1960s, when the Chicago and Detroit transport studies were performed. The early approach is often characterised as a reactive, predict-provide approach in which vehicle and passenger volumes in the main travel corridors were estimated and increases in road and public transport capacities were proposed to accommodate those expected increases for the long-run. During that period the models were also typically single-mode, or unimodal. Since that period transport models have evolved³ into multimode, or multimodal models (from begin 1970s), with a sound theoretical basis, notably the economic theory. Due to increased computing possibilities the transport models have become more disaggregated in detail level and have been applied at larger scales. Data collection techniques and estimation techniques have also improved considerably. Major applications of transport models nowadays are with environmental impact assessments and road pricing measures. The traditional

³Nevertheless, the development of transport planning techniques has been evolutionary rather than revolutionary (Bates, 2001).

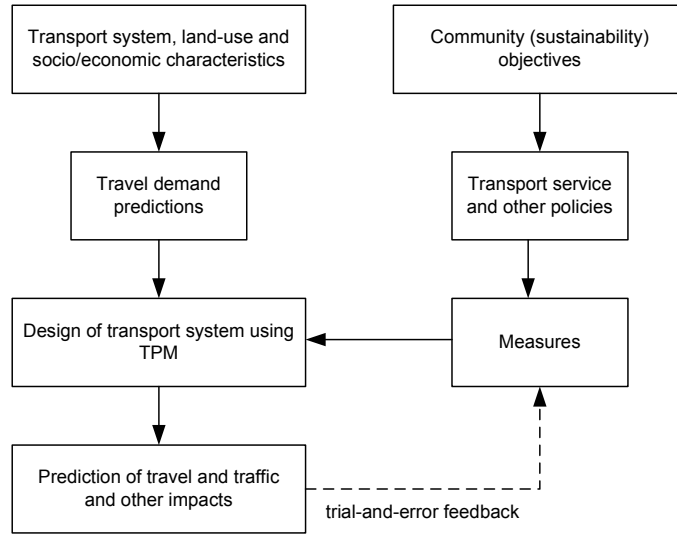


Figure 1.6: Traditional transport modelling process (predict-provide).

transport modelling process is graphically depicted in figure 1.6. Future year projections of travel demand that are based on predicted transport system and traffic characteristics, social-economic development as well as land-use planning, together with some traffic and travel related measures, which are derived from some transport policy objective (possibly based on explicit community objectives), form the basis for the design of the transport system using a transport planning model. At best there is a predict-provide feedback (through trial-and-error) from the predicted impact of transport and traffic figures back to the design of the transport system, in order to try to relief some of the traffic or other (negative) impact.

The so called Transport Planning Model (TPM) or traditional transport model formed ever since the central part of the transport planning process (see figure 1.4). In transport modelling, transport systems are depicted as networks $\mathcal{G}(\mathcal{N}, \mathcal{L})$, involving nodes $n \in \mathcal{N}$ (node-set \mathcal{N} representing cities, zones, possibly also intersections etceteras), which are joint together by capacity restrained links $l \in \mathcal{L}$ (link-set \mathcal{L} representing roads, railways, etceteras). Each zone is represented with a zonal centroid from the centroid-set: $\mathcal{Z} = (\mathcal{I}, \mathcal{J}) \subset \mathcal{N}$, which consists of origins $i \in \mathcal{I}$, and destinations $j \in \mathcal{J}$. In every centroid all trip origins and destinations are located, which generate the traffic flows along the links between the zones in the network. In addition, physical, demographic and social-economic variables useful to define the system of activities are also indicated in these centroids. Furthermore, a mode-set for vehicles $m \in \mathcal{M}$ as well as origin-destination specific route-sets for routes $r \in \mathcal{R}_{ij}$ are defined. In addition, $\mathcal{R} \supset \mathcal{R}_{ij}$ is the set of all routes, \mathcal{K} the set with population segments k and \mathcal{P} the set with pollutants p .

Conventionally the traditional transport model, also four-step model, is divided into four sequentially linked sub models:

Trip generation, which is the number of trips associated with a zone at the node end and consists of trips produced and trips attracted to that zone;

Trip distribution, which is the allocation of trips between each pair of zones in the study area, thus producing an origin - destination (OD) trip table;

Modal split, which determines the number of trips by each mode of transport between each pair of zones;

Trip assignment, which allocates all trips by origin and destination zone to the actual links that comprise the road network. Separate allocations normally take place for each mode. Trips are usually converted to vehicle trips using an average vehicle occupancy rate. Hence, this sub model is better named ‘traffic assignment’.

The first three sub models are concerned with the calculation of travel demand (in person trips) and the fourth sub model with the interaction with the transport system (to reveal vehicle trips). Bates (2001) gives a thorough description on the first three sub models, which forms the basis for the model summary presented here. The modelling of travel demand implies a procedure for predicting what travel decisions people would wish to make, given the generalised cost of all alternatives available to them. The decisions include choice of route (traffic assignment), mode (modal split), destination (trip distribution), and frequency (trip generation). Often, the choice of time of travel is also added⁴. These choices can be linked together using choice hierarchies, implementing discrete-choice theory (for an extensive discussion on this topic the reader is referred to amongst others Ben-Akiva and Lerman (1985)). Lower level choices are made conditional on higher choices in a theoretically consistent way. A possible, idealistic, structure is shown in figure 1.7. That is, most transport models around have a sequential structure, not having this possibility of feedback. If this feedback is allowed, Manheim’s demand-supply equilibrium could be obtained.

The first sub model, the trip generation model, or trip frequency model can be subdivided into a trip production and trip attraction model. In general, the first one can be estimated quite accurately, using the general trip production model:

$$T_{i|k} = f(x_{1i|k}, x_{2i|k}, \dots, x_{ni|k}; [c_{i|k}^*]) = f(\vec{x}_{i|k}; [c_{i|k}^*]), \quad (1.2)$$

where k is a population segmentation (usually trip purpose and social-economic background specific), i is the origin zone in the study-area, while the vector: $\vec{x}_{i|k} = x_{1i|k}, x_{2i|k}, \dots, x_{ni|k}$, represents n social-economic characteristics for po-

⁴The time-of-day sub model is not considered in this thesis.

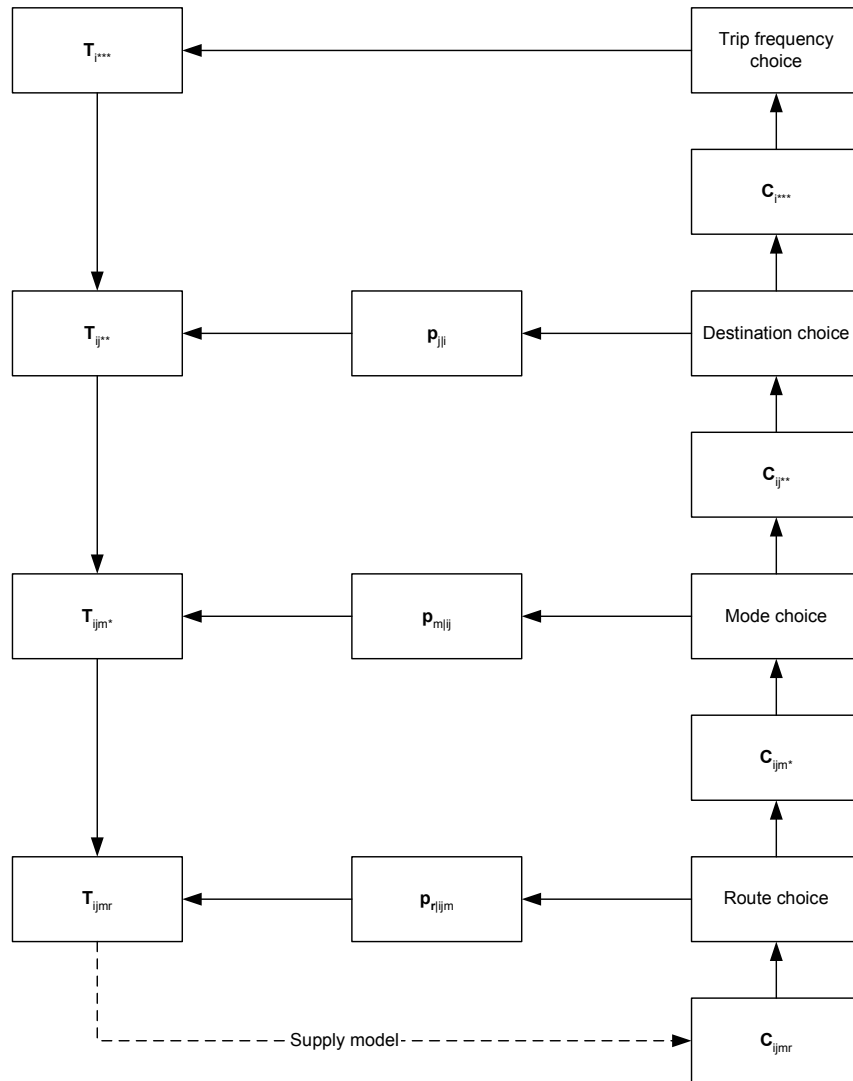


Figure 1.7: The traditional transport model, based on Bates (2001).

pulation segmentation k in zone i , whereas $c_{i|k}^*$ is the generalised cost, or composite cost of travelling from the origin zone i . Bates (2001) deliberately puts these costs of travelling between brackets since it is hardly used in practice; trip production is usually treated as being dependent on exogenous variables to the model only. This is done, even though it is very conceivable that the level of trip making is influenced by the transport system, hence introducing the concept of accessibility. The trip production equation is used to obtain the total number of trip-ends in the zones of study. Most trip production models are household or person-based, implying that a zonal aggregation has to be performed, using

information on the total number of households per segmentation k in zone i , $H_{i|k}$, that is:

$$T_i = a_i^0 + \sum_k H_{i|k} f(\vec{x}_{i|k}; [c_{i|k}^*]), \quad (1.3)$$

with a_i^0 an intercept usually resulting from the model estimation, e.g. applying techniques of multiple linear regression (see paragraph B.1).

Trip attraction models for T_j can have a similar structure as in equation (1.2), where the explaining variables for the different attractions in zone j are related to the type of land-uses attracting produced trips. In general, it appears rather complicated to find these figures. Ideally the trip attraction rates, like the trip production rates, are also dependent on the generalised cost between the zones of production and attraction.

Since trip productions can be more accurately determined, trip attractions are usually balanced (in total made equal) to the total number of trip productions. Therefore, the trip attraction model merely serves as a distributor of trips over the attractors. To balance the total number of trip productions and trip attractions a balancing factor f is applied to all trip attractions T'_j to obtain the balanced T_j , which is:

$$f = \frac{T}{\sum_{j \in \mathcal{J}} T'_j}, \quad (1.4)$$

with the total number of trips being: $T = \sum_{i \in \mathcal{I}} T_i$.

In the second sub model, i.e. the trip distribution model, an OD - table, matrix T_{ij} , is constructed, relating the number of trips in the matrix cell (i, j) to:

1. the characteristics of the origin/production zone i ;
2. the characteristics of the destination/attraction zone j ;
3. the characteristics of the generalised cost of travel, between zones i and j .

This relation, named the gravity model after its apparent analogy with the Newtonian law of gravitation, has the general form:

$$T_{ij} = \mu Q_i X_j f(c_{ij}), \quad (1.5)$$

where Q_i is the production potential of zone i , X_j the attraction potential for zone j , μ the 'gravity' constant⁵ and $f(c_{ij})$ is a distribution function⁶, in

⁵Parameter μ is interpreted here as a measure of average trip intensity in an area, being the number of travellers P divided by the number of and variability in trip alternatives k , i.e.: $\mu = \frac{P}{k}$ (Bovy and Van der Zijpp, 1999).

⁶Also called deterrence function, or impedance function.

its early form depicted as $f(d_{ij})$, hence only considering the travel distance between zones i and j indicated as d_{ij} :

$$f(d_{ij}) = \frac{1}{d_{ij}^2}, \quad (1.6)$$

but nowadays often used in an exponential form with parameter λ , relating the distribution to the generalised or composite costs of travel between zones i and j , c_{ij} :

$$f(c_{ij}) = \exp(-\lambda c_{ij}). \quad (1.7)$$

The exact calculation of T_{ij} usually assumes that trip productions and/or trip attractions from the trip generation sub model are known. In the production constrained trip distribution model, the number of trip productions T_i are imposed as a set of constraints: $\sum_j T_{ij} = T_i$, on the general trip distribution model, after some calculations giving:

$$T_{ij} = T_i \frac{X_j f(c_{ij})}{\sum_{j' \in \mathcal{J}} X_{j'} f(c_{ij'})}. \quad (1.8)$$

Hence, the production constrained trip distribution model is a proportional model that splits the known trip production numbers in proportion to the attraction potential X_j . Similarly, an attraction constrained trip distribution model can be constructed.

If both the number of trip productions and trip attractions are known the calculation of T_{ij} is usually performed as an iterative process known as bi-proportional fitting, or the *Furness* method, where two sets of constraints are given, namely the numbers of arrivals: $\sum_i T_{ij} = T_j$, and departures: $\sum_j T_{ij} = T_i$, as well as two balancing parameters a_i and b_j , changing equation (1.5) in:

$$T_{ij} = a_i T_i b_j T_j f(c_{ij}). \quad (1.9)$$

Obviously, the matrix-total is: $\sum_i \sum_j T_{ij} = T$.

In figure 1.7 this destination-choice model is depicted as a conditional probability:

$$p_{j|i:k} = f(c_{ij|k}^*, c_{i\{j\}|k} : \vec{x}_{i|k}, \vec{z}_j), \quad (1.10)$$

where, as before, k is a segmentation of the population, i and j again the origin and destination zone, $p_{j|i:k}$ the proportion of all travellers of type k in

zone i , who travel to zone j , $c_{ij|k}^*$ the composite cost of travel between zones i and j , and $c_{i\{j\}|k}$ the associate cost of travel to all zones, with $\{j\}$ the set of destination zones \mathcal{J} , and: $\vec{x}_{i|k} = x_{1i|k}, x_{2i|k}, \dots, x_{ni|k}$, the vector of n social-economic characteristics for segmentation k , while: $\vec{z}_j = z_{1j}, z_{2j}, \dots, z_{n'j}$, is a vector of n' zonal characteristics. The reader is referred to Ortúzar and Willumsen (2001) for a complete derivation of the gravity model.

Similarly, the third sub model, i.e. the mode choice model, can be formulated as a conditional probability:

$$p_{m|ij:k} = f(c_{ijm|k}^*, c_{ij\{m\}|k}), \quad (1.11)$$

where, as before, k is a segmentation of the population, i and j again the origin and destination zone, m the mode, $p_{m|ij:k}$ the proportion of all travellers from population segmentation k moving between zone i and zone j who use mode m , $c_{ijm|k}^*$ the composite cost of travel between zones i and j by mode m , and $c_{ij\{m\}|k}$ the associate cost of travel by all modes, with $\{m\}$ the set of modes \mathcal{M} being considered. The main sources of variation in the mode choice models, used in practice, according to Bates (2001) are:

1. the number and type of modes actually distinguished;
2. the detail of the generalised cost functions $c_{ijm|k}^*$.

Mode choice models initially were bi-modal, requiring a sigmoidal curve whereby the probability of choosing the mode vanishes when its costs vary greatly in excess of the costs of the other mode, but which allows reasonable sensitivity when the costs are comparable. For multimodal models, commonly used nowadays, the discrete-choice or multinomial logit (MNL) formulation of equation (1.11) is used:

$$p_{m|ij:k} = \frac{\exp(-\lambda_{1|k} f(c_{ijm|k}^*))}{\sum_{m' \in \mathcal{M}} \exp(-\lambda_{1|k} f(c_{ijm'|k}^*))}, \quad (1.12)$$

with $\lambda_{1|k}$ the ‘spread’ parameter or scale parameter reflecting the degree of substitutability between the modes m , in other words the sensitivity of choice of mode to changes in generalised cost, or utility $c_{ijm|k}^*$. Often, $\lambda_{1|k}$ is chosen to be unity, as in the Maximum Likelihood Estimation (see appendix B.2) of the model, it gets integrated in the parameters for the utility function $f(c_{ijm|k}^*)$, denoted as utility or disutility $u_{ijm|k}$ in short (see below). The effect of variations in the value of $\lambda_{1|k}$, for the binomial case, is depicted in figure 1.8. At lower values of $\lambda_{1|k}$ changes in decision choice, in this case choosing for alternative 1 with disutility, or travel impedance u_1 , in relation to alternative 2 with disutility u_2 , are smoother.

Utility is considered to be the value that individuals derive from choosing a certain alternative, in other words utility is related to the (relative) attractiveness

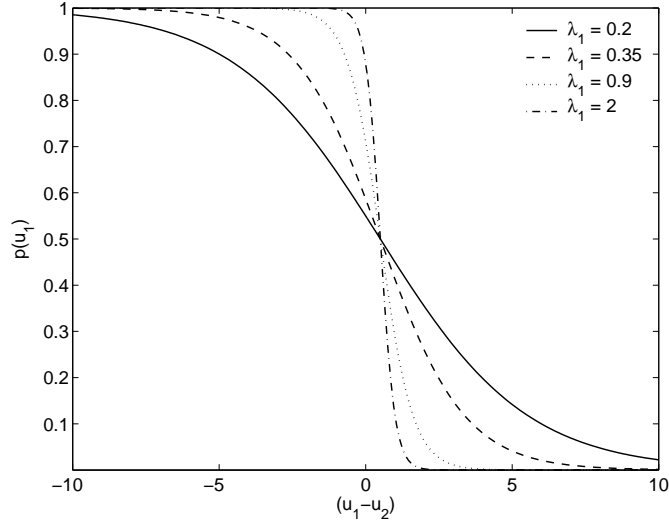


Figure 1.8: Diversion curves for decision choice behaviour.

of alternatives. The net-utility for mode alternative m , for example, experienced by individual or population segmentation k consists then of a measurable, or systematic part $v_{m|k}$ and a random part $\epsilon_{m|k}$, representing particular taste-values, but also observational errors made in the modelling:

$$u_{m|k} = v_{m|k} + \epsilon_{m|k}. \quad (1.13)$$

The measurable part $v_{m|k}$ may be a function of several attributes, like the travel characteristics, travel time and travel cost per mode, with parameters as in equation 1.1. Similar to neoclassical economic theory, the alternative with the highest utility is supposed to be chosen. Therefore, the probability that alternative m is chosen by decision-maker type k within choice-set \mathcal{M} is:

$$p_{m|k} = p \left[u_{m|k} = \arg \max_{m' \in \mathcal{M}} u_{m'|k} \right]. \quad (1.14)$$

The mode-specific travel demand T_{ijm} can now be calculated using the discrete-choice model (1.12):

$$T_{ijm} = \theta_m T_{ij} p_{m|ij:k}, \quad (1.15)$$

where θ_m is the vehicle m occupancy factor.

The travel demand T_{ijm} is in the fourth sub model confronted with the supply model representing the transport system itself. Hence, in the traffic assignment

sub model, the trips T_{ijm} are converted into vehicle trips (go on foot, bicycle, car etceteras) and (iteratively) loaded onto the network of shortest paths with route r specific generalised costs c_{ijmr} between the different origins i and destinations j for the several modes m , allowing or not allowing for congestion to develop. The simplest all-or-nothing traffic assignment model reads:

$$T_{ijmr} = \begin{cases} T_{ijm} & \text{for the minimum cost route } c_{ijmr}, \\ 0 & \text{for all other routes.} \end{cases} \quad (1.16)$$

Here, all trips are assigned to the route with minimum cost, on the basis that these are the routes, travellers would want to use. This situation is only realistic when there is no congestion and when there is only one route with a distinct minimum cost. If there is congestion the assignment model should at least be capacity-restrained, leading to a so called Wardrop's First Principle⁷ or User-Equilibrium (UE) in which all travellers on a certain origin-destination pair perceive equal costs. If congestion is considered, a nonlinear congestion curve (representing the speed-flow relationship), as the supply functions in figure 1.5, is applied. This is often illustrated alongside with the Bureau of Public Roads' (BPR) travel time equation (Bureau of Public Roads, 1964) :

$$\tau_l = \tau_l^0 \left[1.0 + \alpha_1 \left(\frac{V_l}{C_l} \right)^{\beta_1} \right], \quad (1.17)$$

where τ_l is the travel time on link l in the transport network, τ_l^0 the free-flow travel time on link l , V_l and C_l respectively the link volume and link capacity. In addition, α_1 and β_1 are positive-valued parameters. A route then comprises of several links, hence route travel time τ_{ijmr} is depicted as a the summation of individual link travel times that comprise a certain route r between i and j (by mode m), not considering intersection delay, as is sometimes done in this type of models:

$$\tau_{ijmr} = \sum_{l \in r \{ \mathcal{R}_{ij} \}} \tau_l^0 \left[1.0 + \alpha_1 \left(\frac{V_l}{C_l} \right)^{\beta_1} \right]. \quad (1.18)$$

The mode choice and traffic assignment model can also be combined in a simultaneous mode, destination and route choice model, hence combining equations (1.8) and (1.12) into:

$$T_{ijmr|k} = \theta_m T_i \frac{X_j \exp(-\lambda_{1|k} f(c_{ijmr|k}))}{\sum_{j' \in \mathcal{J}} \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij'}} X_{j'} \exp(-\lambda_{1|k} f(c_{ij'm'r'|k}))}, \quad (1.19)$$

⁷In Wardrop's Second Principle, the average trip time of users is minimal, implying that the total cost in the transport network is minimal, which is also called a System-Optimum.

where the total utility $f(c_{ijmr|k})$ of a destination-mode-route choice combination for a given origin, is expressed as a function of individual measurable choice-utility components v and accompanying random elements ϵ (compare equation (1.13)), noting that utility directly consists of time and cost elements c :

$$f(c_{ijmr|k}) \equiv u_{ijmr|k} = v_{ij|k} + v_{ijm|k} + v_{ijmr|k} + \epsilon_{ij|k} + \epsilon_{ijm|k} + \epsilon_{ijmr|k}. \quad (1.20)$$

This formulation only holds when the individual choice-utility components are assumed to be *non-correlated* or independent, and parameter $\lambda_{1|k}$ is unique for the combined choice; see Oppenheim (1995) for a discussion on such formulation. A multinomial logit model then results, in which the argument is the conditional utility (based on travel time $\tau_{ijmr|k}$ and travel costs $\kappa_{ijmr|k}$) for a given simultaneous origin i , destination j , mode m and route r alternative, alike equation (1.1), i.e.:

$$u_{ijmr|k} = \alpha \tau_{ijmr|k} + \beta \kappa_{ijmr|k}. \quad (1.21)$$

From this simultaneous mode, destination, route choice model or the standard traffic assignment sub model, the vehicle link flows q_{lm}^v and link travel times τ_l , hence link speeds s_l , can be obtained through link-route incidence.

Several forms of *ex-post* traffic impact analysis can now be performed using the information revealed in the different sub models, ranging from selected link analysis, for example to reveal congestion levels, to different types of environmental impact studies by applying environmental models, thus revealing data on energy consumption, traffic emissions and noise pollution. An example of determining total traffic emissions E^p for transport pollutant $p \in \mathcal{P}$ in a transport network is adapted from Zietsman (2000), i.e.:

$$E^p = \sum_{l \in \mathcal{L}} \sum_{m \in \mathcal{M}} q_{lm}^v R_{\bar{s}_l mp}^* d_l, \quad (1.22)$$

with $R_{\bar{s}_l mp}^*$ the composite emission rate for pollutant type p and mode type m , at the average link speed \bar{s}_l ; d_l is the length of link l . Similarly, the noise pollution can be calculated per link in the transport network. The equivalent continuous sound level L_l^{eq} measured in decibel [dB] can be obtained using the following equation, also obtained from Zietsman (2000):

$$L_l^{\text{eq}} = 10 \log_{10} \left[\frac{\phi}{15} \sum_m (q_{lm}^v) B \left(\frac{15}{D_E} \right)^{1+v} \right], \quad (1.23)$$

where ϕ is the equivalent subtending angle, v a land dampening factor, D_E an equivalent lane distance and B a function of total traffic volume, mode specific traffic volumes and associated mode specific mean speed, or overall mean speed.

More specific exogenous data on the local surrounding area is sometimes used to reveal more detailed noise and traffic pollutant emissions. An example can be found in the PROMIL model that is used in The Netherlands in combination with transport models and Geographical Information System (GIS) data on the area (Goudappel Coffeng, 2003).

In a similar fashion other effects and impacts can be obtained, like fuel consumption, total number of kilometres travelled in the transport network⁸, total number of passenger trips made and (more difficult) traffic unsafety. However, it should be realised that all these impact models don't feature any feedback to the sub models of the transport model.

1.2.3 Transport modelling alternatives

Despite the fact that limitations of the conventional transport model, in particular the four-step transport model described above, are well known by transport professionals it has been used worldwide ever since the 1960s. Some of the limitations often posed are summarised from Banister (2002):

- The positivistic approach that is data driven and makes no attempt at understanding the real mechanisms of people's travel behaviour. Travel is a derived demand and can be quantified using empirical relationships;
- The sequential decision-making process, whereas there is clear evidence that (travel) decisions are made simultaneously. The feedback loops that may describe a more realistic decision-making process, as in figure 1.7, are seldom used;
- The aggregated character of the models, e.g. zonal aggregation, ignore patterns and uncertainty in behaviour;
- The interactions between land-use and transport through land-use variables, as well as social-economic variables like employment and population are conventionally modelled as exogenous variables;
- The structure of the four-step model makes it difficult to include unconventional or radical policy alternatives;
- Most transport models are static and are calibrated on one set of cross-sectional data only. Hence, coefficients are assumed to be stable over time;
- Not all significant variables are specified in the model, or strong assumptions have been made about them;
- Variability of travel over time is ignored. The time unit of analysis is often the morning peak period.

⁸Often the expression Vehicle Mileage Travelled (VMT) is used.

These limitations, causing a lot of criticism over the past 20 years, are, unfortunately, still relevant today and reflect the in-built resistance to or impossibility of radical change by transport planners. To respond to some of the shortcomings of the traditional transport models in the 1970's and 1980's researchers have been trying to improve the existing approach, most importantly including the discrete-choice models, which lead to:

- The development of land-use transport models, in which the variable land-use is endogenised. The complexity of input data and the choice models in this case, however, is even more demanding than the traditional transport model;
- The development of disaggregate behavioural models that model the individual choice process on a micro level, being person-based or household-based, instead of the aggregate behavioural assumptions that were made in the earlier versions of the traditional transport model;
- The development of disaggregate utility models to better model the choice behaviour of individuals confronted with several options (assuming full information on the alternatives);
- The development of activity-based models that explicitly model travel as a derived demand from activities undertaken instead of the direct demand assumed in the traditional transport model.

In the 1990's and 2000's the basic framework for transport analysis remained as it was in the 1960's, though some other developments have taken place, i.e.:

- A shift in focus from the long-term strategic to more medium-term tactical/operational models, also called Integrated Transport Studies (ITS);
- A focus on evaluation methods as Cost Benefit Analysis (CBA) as well as Environmental Impact Assessment (EIA), coupled with the traditional transport model;
- Stated preference (SP) and contingency valuation methods, to reveal preferences for and the value of, not (yet) existing, choice alternatives, for use in the discrete-choice model;
- Dynamic analysis to reveal more information on long-term elasticities in behaviour of people;
- Large-scale land use and transport demand studies, based on large datasets, mainly for analysing and monitoring system performance.

In this research it is by no means tried to resolve all mentioned limitations and shortcomings to the traditional transport model, because it would be an unfeasible exercise as many researchers have been trying to do so, with various result. Instead, some major shortcomings that are considered to be indisputably related to the topic of this thesis, i.e. sustainable development of urban transport systems, are dealt with. These are discussed in next paragraph.

1.3 Strategic transport planning

In view of the transport realities described before, the theory of sustainable development and sustainable urban transport development - further developed in chapter 2 - and above mentioned critiques to conventional transport system analysis, a conceptual analytical framework and an accompanying dynamic transport model for sustainable urban transport development are proposed in this thesis. There are three main reasons for doing this. First, a directly applicable analytical conceptual framework for sustainable urban transport development, focussing on possible application within (strategic) transport planning and modelling, is not yet existing. Second, the requirements derived from the conceptualisation of sustainable urban transport development are not yet internalised in the traditional transport planning model, partly because the analytical framework is not existing, partly because it is not a trivial exercise to do so. Third, conventional transport models can hardly produce useful recommendations to decision-makers if they are not founded on the understanding of the continuously changing behaviour of its users (the trip makers), the performance of the transport system itself as well as the complex and interacting objectives by its decision-makers.

The internalisation of requirements for sustainable development in transport planning models necessitates a couple of fundamental changes (in arbitrary order) to the existing modelling; First, the omission of feedback to higher-level sub models, mainly because of its complexity should be tackled; in particular in relation to the inelasticity of the trip frequency model to changes in transport system performance. Second, the static character of the transport model is deemed unsuitable, because sustainable development is a dynamic process by itself, whereas the different processes of generating travel demand and changing infrastructure supply require different time scales, hence urging the use of dynamic models, particularly because of the medium to long-term character of the concept of sustainable development. Third, sustainability demands the inclusion of externalities like traffic pollution and their system limits, which is often not the case in current modelling techniques. Fourth, because of the inherent complexity of transport system analysis and its (dynamic) equilibrium, as previously indicated in figure 1.3, it is believed that transport measures taken to reach a certain policy objective, as well as the policy objective itself, should also be endogenised in the process, hence seeking transport system optima for the considered policy objective instead of using a predict-provide (through trial-and-error) feature. Hence, the model is objective-led, featuring a provide-predict structure.

Even, though, particularly for use in medium to long term strategic planning, the basic building blocks in the traditional modelling by itself are believed to be adequate, above mentioned fundamental problems should be resolved in order to make these models suitable for the task of deriving transport policies aimed at sustainable urban transport development. Hence, this thesis presents a dynamic modelling based on the traditional modelling framework that can

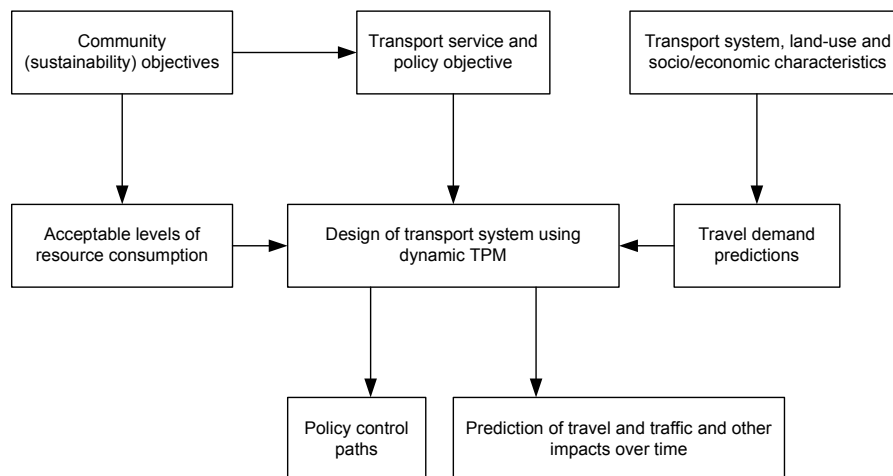


Figure 1.9: Alternative transport modelling process (provide-predict).

optimise transport systems on the basis of a certain transport policy objective, while restricted by sustainability requirements, over time. This can then be used for strategic transport planning, as it produces medium to long term dynamic travel and traffic figures as well as the policy control paths that optimise the chosen sustainable transport policy objective for the transport system under study. This implies that the traditional predict-provide transport modelling process - which can be typified as following a: ‘What if?’ strategy - from figure 1.6, changes into the provide-predict sustainable transport modelling process - typified as following a: ‘How to?’ strategy - as is depicted in figure 1.9.

Strategic transport planning has its roots in long term planning of transport infrastructure and is based on the assumption that decision-makers are responsible for defining objectives and hence decide on a medium to long term transport strategy including sets of transport measures. Strategic transport planning, which is based on the same principles as ‘normal’ transport planning, is therefore typically broad and indicative, in the sense that further detailing of strategies (in time and over time) is required. Such an approach is followed here also.

Hence, the type and detail of decisions that are considered here, can be typified as tactical to strategic. In particular, because this is in line with the medium to long-term perspective that goes with sustainable development. This is well described in Ortúzar and Willumsen (2001), who assume a trade-off between the time horizon of the transport planning decision and the level of detail. At a strategic level, analysis and choices have major system-wide and long-term impacts, and usually involve resource allocation and network design. At a tactical level a somewhat more detailed perspective is chosen, involving questions like

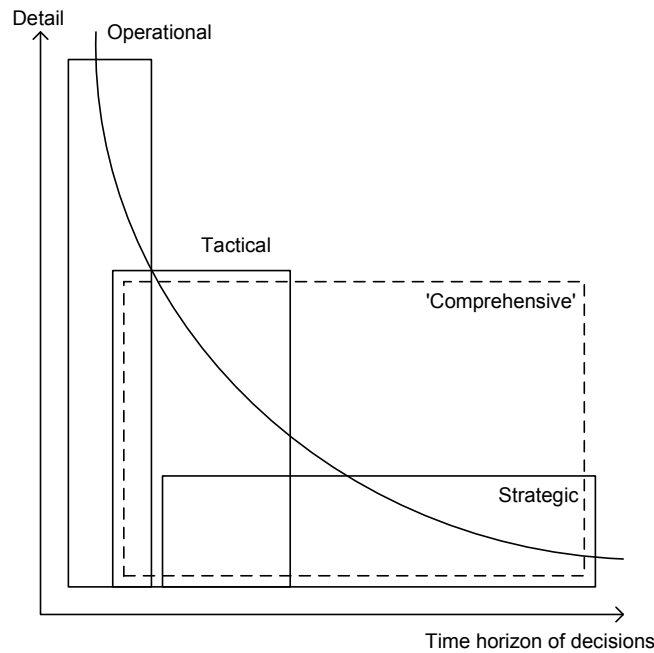


Figure 1.10: Trade-offs in strategic, tactical and operational transport planning, adopted from Lee (1994).

making the best use of facilities and infrastructure. At the short-term a detailed perspective can be adopted, i.e. an operational level of analysis that may include detailed capacity analysis at the link level. This trade-off between perspectives and time horizon of decision-making is depicted in figure 1.10, which is adopted from Lee (1994). The level of decision-making focused on in this thesis, can alternatively be considered as comprehensive, following Lee (1994), as it will show to cover both tactical as well as strategic levels of analysis. In addition, for final and detailed information on, for example, local impacts, the operational level of analysis can still be deployed in combination with the strategic level outcomes, by, for example, adopting a hierarchical modelling approach. The strategic level then functions as indicative and direction-giving for the operational and tactical levels as well as provides a first partitioning of alternatives.

Transport planning seen as an optimisation process gives transport professionals at a strategic or comprehensive decision level, ample possibilities of having sets of policy and engineering measures defined on the basis of a commitment to a certain (sustainable) transport policy objective; yet still complying with the behavioural mechanisms known from traditional transport models. Optimal transport strategies may thus be identified.

The proposed modelling differs from the static frameworks on which it is built,

through its:

Dynamics Transition paths towards a sustainable and developed transport system are revealed;

Optimisation The nature of the problem of designing a sustainable transport planning is that of a constrained optimisation problem. The effects of different transport policy objectives can be studied and compared;

Controls Engineering interventions, such as road construction, road maintenance and public transport priority, as well as pricing measures, like vehicle taxes, parking taxes and bus fares, are the tools, or controls, to the decision-maker;

Constraints Resource constraints and physical constraints to the controls are applied.

Next, a problem statement with accompanying research questions is given.

1.4 Problem statement and research questions

The previous sections gave a nutshell description on the (urban) transport problems and issues that are considered pertinent when discussing (strategic) transport planning and sustainability. Urban transport problems and the requirements for sustainable development are increasingly complex. Current transport systems and transport planning models (used in developing and developed) countries are not necessarily compatible with the requirements of sustainable transport development. Adequate transport systems can only be obtained with use of a new transport planning paradigm and accompanying analytical framework.

A suitable analytical transport planning technique distinguishes itself from traditional planning techniques if it is directly based on a conceptualisation of sustainable development and when it is able to adequately cope with the observed drawbacks with current transport modelling (in particular related to the definition of sustainable development). That is, the current process determines a static equilibrium solution, whereas sustainable transport development by definition is a dynamic phenomenon with several and different coevolving states of its individual systems and users. Moreover, the trip frequency model is considered deficient in its inability to estimate generated or latent demand when transport system performance changes. In addition, sustainable transport development requires a balanced set of planning instruments that reckons with positive as well as negative externalities that come with transport, while complying with a certain transport policy objective, something which is not explicitly considered in current transport modelling.

Definition and description of the implications of adopting a sustainable trans-

port development paradigm as well as the development of an analytical framework and model for sustainable urban transport development are the topic of this thesis. A problem statement and research questions are given below.

The basic problem addressed in this thesis is:

Problem definition: What are the requirements for sustainable urban transport development and what are the implications of these requirements for transport planning, in particular transport modelling?

Internalisation of the concept of sustainable development implies using this concept in the development of new transport systems and plans as well as the management of existing ones. The focus, however, is on the analysis, modelling and forecasting stage of the transport planning process depicted in figure 1.4. The implications of this internalisation on the total transport planning process are, of course, also identified.

The major aim of the research is therefore:

Research aim: To define and describe the requirements for sustainable urban transport development as well as to develop an analytical transport planning method and tool, in which these principles of sustainable development are internalised, and to demonstrate plausibility and feasibility of the ideas and method.

Hence, the research aims at contributing to the achievement of an overall sustainable development in urban areas around the world. Sustainable development, however, goes much further than urban transport sustainability alone. And, also within transport there are many different areas that contribute to sustainability, amongst others freight, maritime transport as well as air transport. However, this thesis will only focus on passenger transport in urban areas. At a geographical scale this research is not limited to cities in developing or developed countries. Both type of cities will somehow be discussed, as the specific manifestation of their urban transport problems may differ, but the underlying transport mechanisms are assumed to be the same.

On a more detailed scale, this study intends to answer the following three research questions:

Research question 1: What are the implications of the notion and definition of sustainable development for urban transport planning?

Research question 2: How should sustainable development be modelled and incorporated in urban transport planning practice?

Research question 3: What typical consequences can be expected and implications be drawn for urban transport planning in the long-term due to internalisation of the concept of sustainable development?

The remainder of this thesis is confined with the answering of these research questions, which are successively addressed throughout the three parts of this thesis.

1.5 Limitations of this research

The research described in this thesis is limited to the study of sustainable development in urban transport planning and modelling; in particular urban passenger transport, i.e. the movement of people in transport networks where destination, mode and route choice behaviour are prevalent. However, it is tried to keep the introductory chapters 1, 2 and 3 general introductions in sustainable development and transport.

Sustainable development is known to be a complex and integrated issue. A full integration will probably never be accomplished. Putting too many requirements on the concept may even lead to failure of achieving anything approaching a sustainable system. Therefore, it is always tried to keep the reader informed where, how and why part of the integration is not established.

The focus in this research is on quantitative aspects of transport planning at a strategic application level. Furthermore, the author's reasoning comes principally from an engineering and mathematical economics point of view.

1.6 Scope and outline of this thesis

This thesis is divided into three main parts. The remainder of this part, part I, deals with the concept of sustainability and sustainable development. In chapter 2, a conceptual framework for sustainable development is given and linked to transport. Some of the many concepts and definitions are given and discussed. A conceptualisation of sustainable transport is then given that is adopted in the remainder of the research. In chapter 3, some more detail is given on the linkage between travel, infrastructure and sustainable urban transport development. Part II deals with modelling sustainable urban transport development. First, in chapter 4, the intended use of optimisation as a tool for sustainable transport policy design is discussed. Followed, in chapter 5, by a description of the dynamic transport model, including some background on mathematical systems theory. Finally, part III deals with an analysis of the implications of using the concept and dynamic transport model for (strategic) transport planning in chapter 6. The intended use of the model in real-life is also discussed. The several consequences that can be expected and implications that can be drawn for urban transport planning in the long-term due to internalisation of the concept of sustainable development are given on the basis of these examples in the concluding chapter 7. This thesis ends in chapter 8 with a short reflexive dialogue on transport sustainability with respect to developing countries.

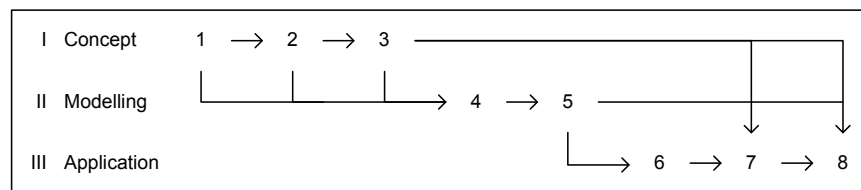


Figure 1.11: Structure of the thesis.

The structure of the thesis, including part and chapter numbers, is given in figure 1.11.

Chapter 2

Sustainable development and transport

2.1 The Limits to Growth

In 1972, also the year that the first United Nations Conference on the Human Environment was held in Stockholm, Sweden, an influential book was published named *The Limits to Growth* by Meadows et al. (1972). It described the prospects for growth in the human population and the global economy for the 20th century. This document created an uproar. The combination of the use of a computer model, and the involvement of The Club of Rome and Massachusetts Institute of Technology (MIT) pronouncing upon pending disaster had an irresistible dramatic appeal.

The three main conclusions in this book were (Meadows et al., 1972):

1. If the present growth trends in world population, industrialisation, pollution, food production, and resource depletion continue unchanged, the limits to growth on this planet will be reached sometime within the next 100 years. The most probable result will be a sudden and uncontrollable decline in both population and industrial capacity;
2. It is possible to alter these growth trends and to establish a condition of ecological and economic stability that is sustainable far into the future. The state of global equilibrium could be designed so that the basic material needs of each person on earth are satisfied and each person has an equal opportunity to realise his or her individual human potential;
3. If the world's people decide to strive for this second outcome rather than the first, the sooner they begin working to attain it, the greater will be their chances of success.

In other words, physical limits to growth exist and, if and only if fully respected, an equitable distribution of well-being is regarded feasible.

In 1984, the World Commission on Environment and Development (WCED) was established with the task of formulating ‘a global agenda for change’, which resulted in 1987 in the publication of the *Our Common Future*, or Brundtland report by WCED (1987)¹. This report investigated the capacity of the earth to support its population and the ways in which human activities were affecting the environment. The Brundtland Commission defined sustainable development as ‘development which meets the needs of the present without compromising the ability of future generations to meet their own needs’.

The two key aspects in this definition and the accompanying Brundtland report are:

1. the realisation of basic needs for all people, in particular those in need;
2. the limits to growth are technical, cultural and social.

These aspects contrast to the limits of growth of Meadows et al. (1972), which are merely environmental and resource availability. Basically, the Brundtland report rejects the physical limits to growth and stresses the belief that equity, growth and environmental sustainability are simultaneously possible.

In 1992 there would be a sequel to the initial *The Limits to Growth* publication, influenced by amongst others critique to the 1972 publication, the public debate on sustainable development that commenced since the appearance of the Brundtland report as well as the second United Nations Conference on Environment and Development (UNCED) in 1992 in Rio de Janeiro, Brazil, where another important document was launched, i.e. *Agenda 21*, in which it is advocated that individual countries should prepare strategies and action plans to implement sustainability^{2,3}. In this sequel *Beyond the Limits* the same authors, Meadows et al. (1992), alter some of the assumptions from 1972, reconfirm some of the conclusions of 1972 and show that some limits of growth have been reached or even surpassed. Some options for sustainability have been narrowed, others have opened up. The three main conclusions of 1972 are in 1992 rewritten as (Meadows et al., 1992):

1. Human use of many essential resources and generation of many kinds of pollutants have already surpassed rates that are physically sustainable. Without significant reductions in material and energy flows, there will be in the coming decades an uncontrolled decline in per capita food output, energy use and industrial production;

¹Named Brundtland report after the chairwoman of the WCED, Gro Harlem Brundtland, who nowadays heads the World Health Organization (WHO).

²On a smaller geographical scale often *Local Agenda 21*'s are introduced to indicate strategies and action plans for sustainability by regions, cities or small communities.

³Unlike the Meadow's reports as well as the Brundtland report, Agenda 21 specifically mentions traffic and transport strategies as tools for achieving a sustainable society. See for example Chapter 7: Promoting sustainable human settlement development, Programme area (e): Promoting sustainable energy and transport systems in human settlements, in UNCED (1992).

2. This decline is not inevitable. To avoid it two changes are necessary. The first is a comprehensive revision of policies and practices that perpetuate growth in material consumption and in population. The second is a rapid, drastic increase in the efficiency with which materials and energy are used;
3. A sustainable society is still technically and economically possible. It could be much more desirable than a society that tries to solve its problems by constant expansion. The transition to a sustainable society requires a careful balance between long-term and short-term goals and an emphasis on sufficiency, equity and quality of life rather than on quantity and output. It requires more than productivity and more than technology; it also requires maturity, compassion and wisdom.

In other words, human activities are affecting the environment, hence technical, cultural and social limits are imposed, not just environmental resource limits as was the case in the 1972 episode.

Again these conclusions were based on (by that time) extensive computer simulations of different global scenarios, applying system dynamics models. Because of the global character, transport was obviously not an explicit part of the models. Figure 2.1 (next page) shows three scenarios that have been calculated in Meadows et al. (1972) as well as Meadows et al. (1992). Scenario (a) shows the ‘standard’ run from *The Limits to Growth*, in which the computer model WORLD3 is run that models stocks as population, industrial capital, pollution and cultivated land. Those stocks change through flows such as births and deaths, investment, depreciation, pollution generation and pollution assimilation. The causal relationships between them are highly nonlinear. Population and industry output grow until a combination of environmental and natural resource constraints eliminates the capacity of the capital sector to sustain investment. In scenario (b) a doubled stock of exploitable resources is assumed. Industrial outputs may grow a mere 20 years longer, but otherwise nothing is gained and the fall is all the steeper. The general behaviour of the model is, like in scenario (a), overshoot and collapse. The system is drawing resources or emitting pollutants at an unsustainable rate, but the stresses on the support system are not yet strong enough to reduce the rates of withdrawal or emission. This overshoot comes from the delays in feedback - from the fact that decision makers in the system do not get, or believe, or act upon information that limits have been exceeded until long after they have been exceeded (Meadows et al., 1992). In that case overshoot may even convert to collapse. Scenario (c) presupposes that sustainability policies, including those directed at population growth, were implemented as early as 1975. It shows that society reaches its desired level of industrial output per person and is able to maintain it and support its improving technologies with no problems. Unfortunately, doing the same now, doesn’t prevent heavy turbulence to occur. Hence, the title *Beyond the Limits ...*

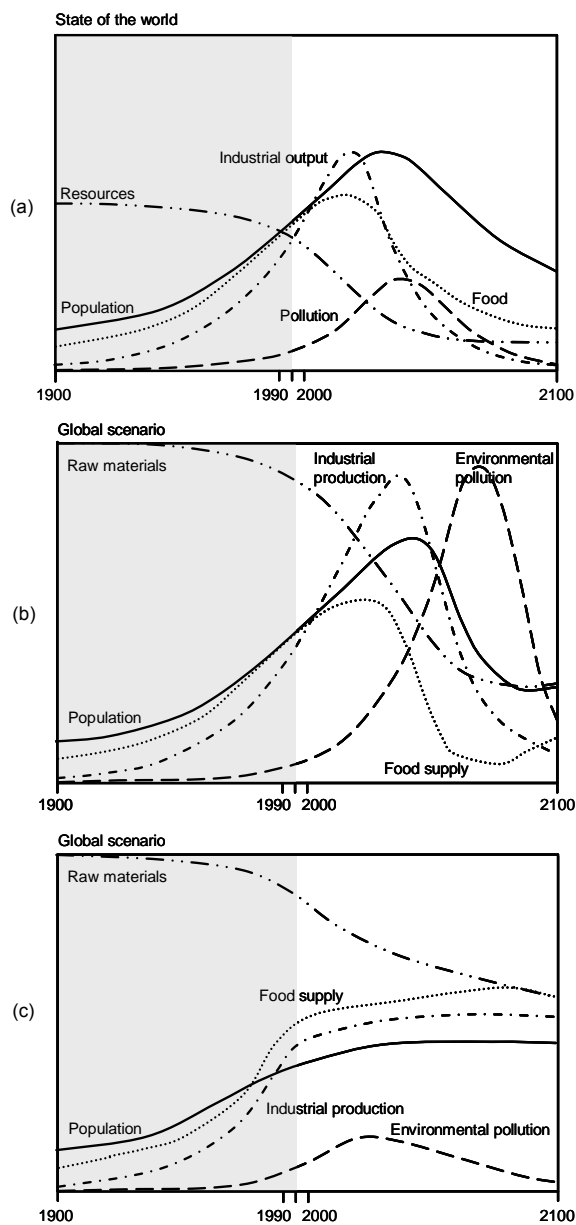


Figure 2.1: (a) The standard run from *The Limits to Growth*. Population and industry output grow until a combination of environmental and natural resource constraints eliminate the capacity of the capital sector to sustain investment; (b) Assuming a doubled stock of exploitable resources, industrial outputs may grow a mere 20 years longer, but otherwise nothing is gained and the fall is all the steeper; (c) The sustainability scenario presupposes that strict sustainability policies including on population, were introduced as early as 1975 (Meadows et al., 1992; Meadows et al., 1972).

Ten years after Rio de Janeiro, in 2002, progress on the ideas and targets set by the Brundtland Commission have been discussed in Johannesburg, South Africa. Although this summit has been criticised for failing to set and renew concrete targets, renewed commitment for action, in particular strengthening the role of the private sector, in achieving sustainability and sustainable development were obtained.

2.2 Sustainability and sustainable development

Sustainable development strives for an optimal balance between economic, social and ecological objectives. Sustainability thinking reflects concerns about long-term risks of current resource consumption, keeping in mind the goals of intergenerational equity. This is posed clearly in the more detailed definition on sustainable development, which is also defined by WCED (1987): ‘in essence sustainable development is not a fixed state of harmony, but rather a process of change in which the exploitation of resources, the direction of investments, the orientation of technological investment, and institutional change are all in harmony and enhance both current and future potential to meet human needs and aspirations’. The starting point is given by economic and social development that can satisfy human needs and aspirations, but also takes into account the natural resource base and social equity. Unlike both Meadows’ reports, the Brundtland report takes a more positive stand, in stating that the world’s resources are sufficient to meet long-term human needs. A distinction, however, should be maintained between *growth* (increased quantity) and *development* (increased quality) (Daly, 1991). Sufficiency and efficiency are important keywords for development⁴.

There is some confusion about the meaning of sustainable development. This has to do with the various interpretations of the terms *sustainability* and *development*. The underlying philosophy of sustainable development indicates that it involves two main aspects - sustainability *and* development. In addition, the basic idea is to harmonise the two aspects in societal activities. However, the focus of most people and organisations is only on sustainability and is interpreted exclusively in terms of keeping the environment in a notionally ‘good’ condition. Sustainability is then a set of restrictions that must be met by society. If these restrictions, which depend on the physical boundaries of nature and maximum capacity of the system for the various resources consumed, are violated, overshoot and collapse may result as seen in paragraph 2.1. Parallel to this, development is often seen as the process of a continued economic and social change, developing, but also developed countries, undergo. The standard of well-being or social welfare is improved, ideally maximised. Considering both aspects simultaneously, a framework of sustainable development is dynamic and relates the factors that define when a system is sustainable to the maximisation

⁴Some reports to The Club of Rome specifically deal with this increase in resource productivity, notably halving resource use in Van Dieren (1995) and more recent, doubling wealth, while halving resource use in *Factor Four* by Von Weizsäcker et al. (1998).

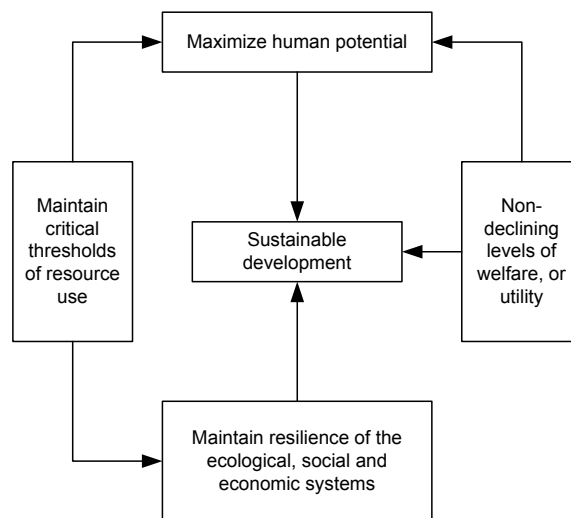


Figure 2.2: A schematic representation of sustainable development, adapted from ESCAP-AITD (2001).

of social welfare, or more general human potential. This framework, read as a dynamic process, rather than as a static objective of optimisation, following ESCAP-AITD (2001), is shown in figure 2.2. Non-declining levels of welfare, or more general non-decreasing utility levels, guarantee that human potential remains constant or is improved over time, while keeping the critical thresholds or capacities of the different resources consumed by the system. The ability of the system to absorb internal and external disturbances to its environment and restore the balance (or equilibrium) is called the *resilience* of the system.

This resilience relates to another important discussion in defining sustainability, i.e. that of *strong* versus *weak* sustainability. Strong sustainability refers to non-decreasing patterns of environmental and resource stocks over time. Weak sustainability allows them to decrease temporarily so long as they return to their initial levels. In this case single stocks may decrease and even be exhausted, so long as the aggregate condition is being satisfied at all times. Non-renewable resources such as stocks of fossil fuels and ores provide a special problem in this discussion. While for renewable resources maintenance of stocks is a legitimate objective, for non-renewables the basic question is how much resources should be left available to future generations (Van den Bergh, 1991). Herman Daly's interpretation of a sustainable society in his book *Steady-State Economics* (Daly, 1991) is an example of a strong approach, in which environment is interpreted as a combination of natural and man-made conditions. According to him a sustainable society is one that satisfies three basic conditions, i.e.:

1. Its rates of use of renewable resources do not exceed their rates of rege-

neration;

2. Its rates of use of non-renewable resources do not exceed the rate at which sustainable renewable substitutes are developed;
3. Its rates of pollution emission do not exceed the assimilative capacity of the environment.

These conditions also reveal that Herman Daly interprets sustainability only in terms of keeping the natural environment in a good state⁵, leaving aside economic and social development.

2.2.1 Three conceptions of sustainable development

Sustainable development can only be reached if sustainability requirements are harmoniously linked to development objectives, as depicted before in figure 2.2. Sustainable development is therefore seen as a bipartite, or multi-directional concept as opposed to a single-directional and dichotomous conception, all three defined by Gudmundsson and Höjer (1996). These fundamentally different intellectual conceptions of sustainable development are shown in figures 2.3(a) to 2.3(c) (next page), and shortly discussed below:

- a. A single-directional conception is a conventional analysis of sustainable development, where all criteria are measured in the same unit. Sustainable development is then more or less just about ‘getting the prices right’, a form of very weak sustainability;
- b. A dichotomous conception is an equally simple version of what could be called ‘eco-centrism’, where sustainability and development are seen as contradictions to each other. In this sustainable development remains an illusion, a form of very strong sustainability;
- c. A multi-directional conception is a conception where sustainability and development represent different dimensions. Sustainability refers to criteria for long-term stability of the social system, relevant for future generations, while development is the perceptible improvement of the quality of human life, of which consumption for the present generation is an important element.

In the multi-directional conception four distinctive states can be observed according to Gudmundsson (2003); only one state being a ‘sustainable development’. Henrik Gudmundsson distinguishes furthermore, first, the states ‘myopic hedonism’ equal to the traditional growth perspective, i.e. an increase in production or output, while not caring (enough) about sustainability. Critical resource capacities might be surpassed. On the contrary there is the second

⁵Although supporters of the steady-state paradigm don’t preclude development, as according to them it finds full expression as the sustainable society evolves (Pearce and Turner, 1990).

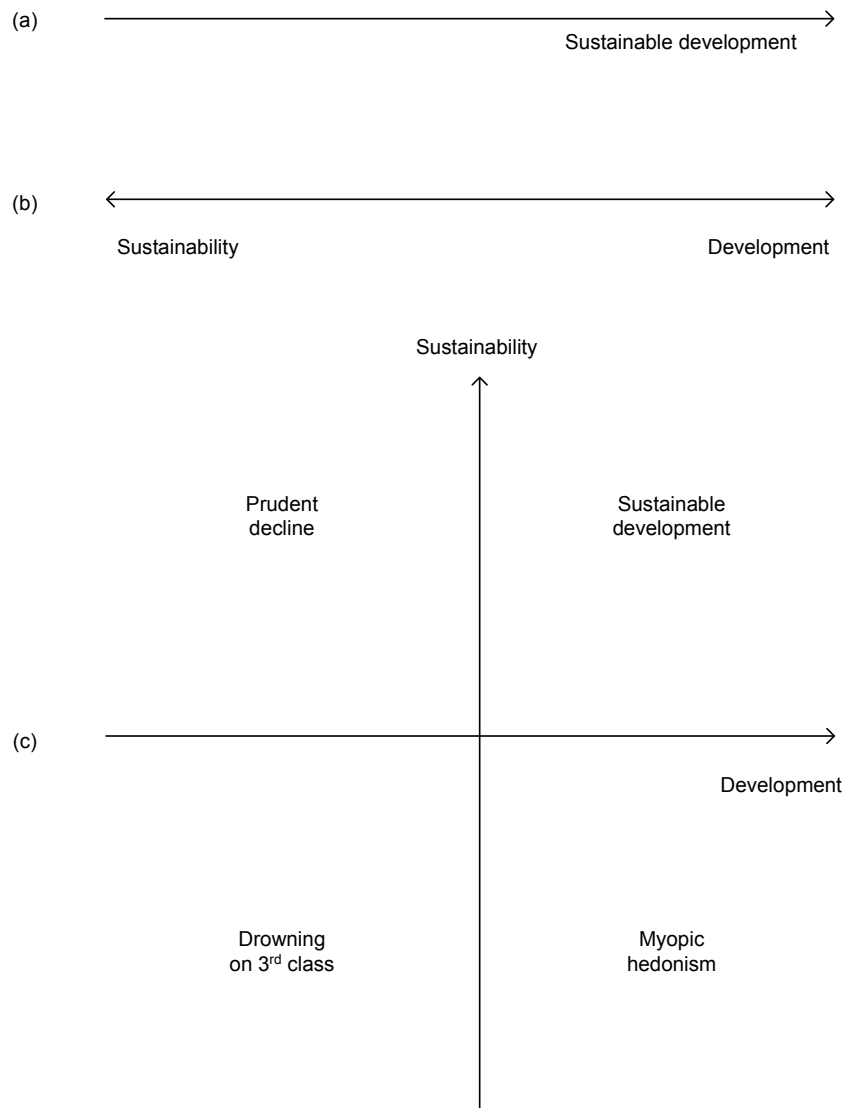


Figure 2.3: Three conceptions of sustainable development: (a) Single-directional concept of sustainable development; (b) Dichotomous concept of sustainable development; (c) Multi-directional concept of sustainable development, after Gudmundsson and Höjer (1996) *ibid.* Gudmundsson (2003).

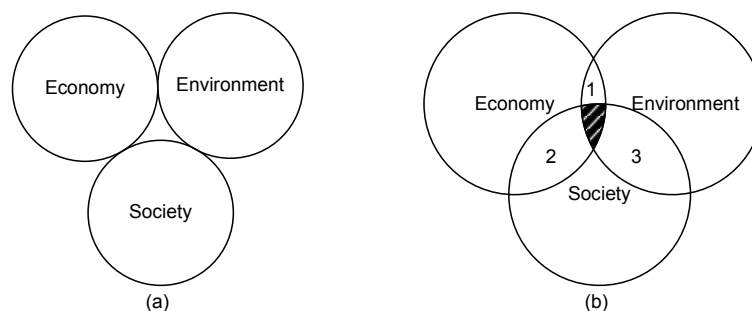


Figure 2.4: Three dimensions of sustainable development: (a) three separate, unrelated parts: economical, environmental and social; (b) three linked parts in sustainable development.

state of ‘prudent decline’, in which the resource capacities still hold, but human potential is not (enough) exploited. The third, obviously the worst case, is one where neither resource capacities are met, nor human potential is exploited, expressively named ‘drowning on 3rd class’. Only in the last case where both the critical capacities of resource use are met as well as human potential is exploited (with non-decreasing general utility levels) there is a ‘real’ sustainable development.

2.2.2 Three dimensions of sustainable development

In a sustainable development context environmental, economic and social sustainability are inseparable on the short and long term. Traditionally they were seen as three separate and unrelated parts, as depicted in figure 2.4(a). In the sustainability approach these three parts are integrated, as can be seen in figure 2.4(b); sustainable development is the highlighted area, where the interaction between economical, environmental and social development is complete. The other overlapping areas can be seen as different paradigms to development, named (1) conservationism, (2) community economic development and (3) deep ecology, as discussed in detail in ICLEI (1996).

Environmental and ecological sustainability deal with maintaining the value of natural systems to provide goods and services, by respecting the maximum capacity of the (local) environment, conserve and recycle resources and reduce wastes. This is done imposing either a weak or strong view on sustainability.

Economic sustainability is concerned with sustaining economic growth by allocating resources, maximising human welfare, expanding markets, internalising costs of externalities, while keeping financial viability. Such externalities exist when the activities of one group of individuals (either consumers or producers) affect the welfare of another group without any payment or compensation (ESCAP-AITD, 2001). Economic sustainability can be linked to financial sustainability, which defines that an economic activity can be performed if it,

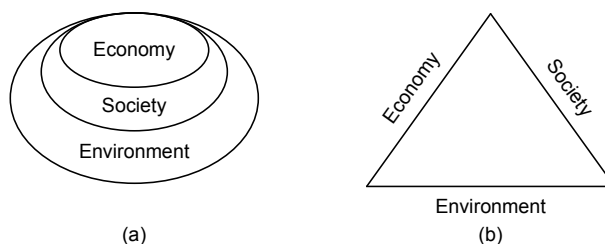


Figure 2.5: Two views on sustainable development: (a) ecological-economics view; (b) World Bank view on sustainable development.

following ESCAP-AITD (2001):

1. attracts sufficient funds to finance the necessary investment and operation;
2. generates sufficient revenue to recover both the operating and capital cost involved;
3. provides the necessary financial incentives to attract and sustain wider participation in such ventures.

Social sustainability is about satisfying basic human needs as being able to (go to) work, attend school, visit healthcare facilities etceteras. These abilities should be equitably distributed amongst people and over time. In other words inter-generational and inter-temporal equity are seen as cornerstones of sustainable development.

The way the three dimensions of sustainable development are interlinked to each other is treated differently by several authors. In an ecological-economics⁶ view to sustainable development, for example, environment and society are seen as prerequisites for economic growth, and hence placed in concentric circles, as in figure 2.5(a). The World Bank, in addition, advocates a more strategic and balanced view on sustainable development, where the three dimensions are displayed in a triangle, as in figure 2.5(b). To be effective any policy must satisfy all three dimensions (World Bank, 1996).

2.2.3 Three levels of discourse

Some authors, notably Becker et al. (1997) and Gudmundsson (2003), emphasise the importance of making a distinction between an *analytical* dimension, a *normative* dimension and a *strategic* or *political* dimension, each defining

⁶Ecological economics is a recently developed field, which sees the economy as a subsystem of a larger finite global ecosystem. Ecological economists question the sustainability of the economy because of its environmental impacts and its material and energy requirements, and also because of the growth of population. (Martinez-Alier, n.d.).

sustainability and sustainable development in a different context. These levels of discourse are closely linked to each other.

At an analytical dimension sustainability is a qualification and/or quantification of states and processes within a continuum of possible states and processes that are rated sustainable. In other words, it is tried to explain how systems like transport systems work, how systems, or their subsystems, interact, and what options exist to control them.

With respect to the normative dimension, sustainability implies the identification of values and the acknowledgement of a hierarchy in these values and how they relate to one another, see for example figure 2.5(a), where in an ecological-economics view the economy depends on society and environment. While societies are possible without a (market) economy, neither can exist without a natural environment. Furthermore, at a normative level it is questioned how to operationalise these values.

Strategically, sustainability requires the identification of different goals and the ways and means of their implementation, the institutional arrangement and the identification of the interests and strategies of possible actors (and conflicts among them). This also means identifying and transforming existing mechanisms of non-sustainability. In other words, the practical implementation of sustainability policy.

From this it should be clear that these dimensions are hierarchically related. Therefore, without first having a thorough understanding of the analytical dimension of sustainable development, the normative and strategic dimensions seem ill-founded. The analytical dimension provides the pre-requisites for a normative discussion, that subsequently provides the foundations for a strategic discussion.

2.2.4 Sustainable urban development

In the previous paragraphs a sustainable and developed system is described as one where human potential is exploited for the foreseeable future without collapse or depletion of the resource base upon which they depend, following *The Limits to Growth* and *Our Common Future* studies. Sustainable systems, like ecological systems, are not steady-state systems, but rather dynamic systems with many feedback loops to provide self-regulation and to keep growth of each part of the system coordinated with the other parts as the system evolves (Replogle, 1991). As the extended Brundtland definition in paragraph 2.2 puts it clearly, sustainable development is therefore a process, not a fixed state.

Summarising, sustainable development can be better understood if the following taxonomy of the concept of sustainable development is made⁷:

⁷This idea of classifying sustainable development as such has been put forward by Henrik Gudmundsson from FLUX Centre for Transport Research at Roskilde University, Denmark

1. Sustainable development consists of two distinct components, i.e. sustainability and development. Sustainable development is therefore best explained as a multi-directional concept;
2. Sustainable development has three dimensions, i.e. economic and financial sustainability, environmental and ecological sustainability as well as social sustainability. These are interlinked and should be in an optimal balance;
3. Sustainable development has three levels of discourse, i.e. analytic, normative and strategic. In particular, in the analytical level it is tried to understand how (sustainable) systems work.

Preceding discussion, found upon both Meadow's and Brundtland reports, is mainly directed at a global, regional or country scale. *Urban* transport, the topic of this thesis, however, may put different demands on the issue of sustainable development, as is well described in Camagni (1998) *ibid.* Camagni et al. (1998).

A greater part of environmental problems originate from the internal operating mechanism of cities, and have their local impact, as air pollution, congestion and noise pollution. However, many effects also exist that have a transborder nature, for example waste water flow, waste disposal, or even a global one, through contribution of traffic and heating emissions to greenhouse effect and global warming. Therefore it sounds plausible to regard cities as starting points for sustainability policies, if consistent with policies elsewhere up to the level of global environmental decision-making (Camagni et al., 1998). Hence, *Local Agenda 21* plans have been introduced in many cities since the 1992 Rio de Janeiro conference.

As low environmental quality can be regarded detrimental to (economic) development of a city, since it generates negative stimuli for human health, education and social welfare, sustainability should be seen in a wider sense than the pure, be it strong, ecological one. In this ecological approach, cities are inherently unsustainable because of their dependence upon the existence of importing resources, many of them non-renewable, and exporting waste. In this research, following this interpretation of urban sustainability by Camagni et al. (1998), environmental quality is enlarged to economic and social environments constituting the city, as the three dimensions of sustainable development depicted before. A sustainable city is then seen as one where these three dimensions interact in such a way as to at least maintain, but preferably increase, the quality and quantity of positive externalities, as accessibility, health, education, social amenities, while at the same time decrease or even bound the external effects caused by this interaction. As Nijkamp and Opschoor (1995) put it (from an economic point of view), 'a sustainable city is a city in which agglomeration economies [pointing at the increasing return in the use of scarce non-renewable

in a Ph.D. course on the topic (Gudmundsson, 2003).

resources as concentration of activities and proximity increase] should possibly be associated with positive environmental externalities and social network externalities, and in which at the same time negative effects stemming from the interaction of the three different dimensions are kept within certain capacity conditions associated with the urban carrying capacity on the urban environmental utilisation space'. Equally, Newman and Kenworthy (1999), define the goal of sustainability in a city as the reduction of the city's use of natural resources and production of wastes, while simultaneously improving its liveability, so that it can better fit within the capacities of local, regional and global ecosystems.

Hence, a sustainable urban transport system should be associated with its positive economic, environmental and social externalities in which at the same time the negative effects stemming from the interaction of the three different dimensions of sustainable development are kept within certain capacity conditions associated with the urban transport carrying capacity. Sustainable urban transport development can now be seen as an area of specialisation within sustainable development. The conceptualisations, taxonomy and urban focus depicted above can be used to further analyse the concept of sustainable urban transport development, as is done in the next paragraph.

2.3 Sustainable urban transport development

Transport improvements undoubtedly promote economic growth and social development by increasing mobility and improving accessibility to people, resources and markets, concepts that are discussed in some more detail in chapter 3. There have, however, been some concerns about the effect of (improved) transport systems on sustainable development. Several dimensions to the impact of traffic and transport and its associated costs that it imposes can be distinguished alike the general framework of sustainable development, i.e. economic, social and environmental, for both current and future generations. A sustainable urban transport development can only be achieved if full account is taken of all these aspects. This paragraph discusses the characteristics of sustainable urban transport development.

The multi-directional conception of sustainability as depicted before in figure 2.3(c) also applies to transport. Transport enables development needs of individuals, companies and societies to be met through the provision of (basic) transport services. Here, transport services consist of mobility (quantity of transport opportunities offered) and accessibility (quality of access between origins and destinations) options provided for. Sustainability requires this to be achieved in a manner consistent with public health and ecosystem capacity, as to promote equity within and between successive generations.

As sustainable development is claimed to integrate economic, social and environmental sustainability, so is sustainable transport development. With respect

to economic sustainability, transport systems should (following The Centre for Sustainable Transportation, Canada (CST, 1997)):

- provide cost-effective transport services and infrastructure capacity;
- be financially affordable (to each generation);
- support vibrant, sustainable economic activity.

This implies that affordable financial resources should be allocated to parts of the transport system, that create or improve transport services in support of human potential in the area.

With respect to social sustainability, transport systems should:

- meet basic human needs for health, comfort, convenience and safety;
- allow and support development of communities, and provide for a reasonable choice of transport services.

This refers generally to an equitable improvement of standards of living and quality of life, by making transport available to all members of society.

With respect to environmental sustainability, transport systems should:

- make use of land in a way that has little or no impact on the integrity of ecosystems;
- use energy sources that are essentially renewable or inexhaustible;
- produce no more emissions and waste than the transport system's carrying capacity;
- produce no more noise than an acceptable threshold of noise pollution.

Environmental effects differ in the locality of their impact. Traffic related air pollution⁸, due to amongst others carbon monoxide (CO), nitrogen oxides (NO_x), particulate matter (PM), volatile organic compound (VOC), and noise have local impacts, whereas carbon dioxide (CO₂), for example, has a serious global environmental impact, being a greenhouse gas (GHG). To control emissions thresholds for urban environmental carrying capacity by pollutant or represented by a proxy for one dominant pollutant should be introduced. This capacity might for example be based on conventions as European Union emission standards or the notorious *Kyoto Protocol* where, back in 1997, countries agreed on emission targets under the *Framework Convention on Climate Change*. Under this protocol, which is effective per February 2005, developed countries promise to stabilise their emissions to 1990 levels in 2000, and reduce Greenhouse Gas emission to 95% of 1990 levels by 2008 - 2012.

⁸Appendix A describes the different pollutants in some detail.

2.3.1 A framework for sustainable urban transport development

Preceding discussions have shown that sustainable transport development is basically a dynamic process of harmonisation of sustainability and transport development requirements. It is therefore that a sustainable and developed transport system is postulated to be: ‘a transport system that meets the people’s transport related needs in terms of mobility, accessibility and safety, within limits of available or affordable environmental, financial and social resource capacities⁹’. However, there are many other interpretations to sustainable transport. Some of the characteristics and problems of the common interpretations have been highlighted in Akinyemi and Zuidgeest (2000) *ibid.* Zuidgeest et al. (2000). The current focus and most of the debate have been generally limited to (a) the issues that need to be resolved for transport to conform to the principles of sustainable development and (b) the required policy paths. The idea has not yet (effectively) reached planning and operational practices and actions in transport infrastructure planning and management. Consequently, two major needs can be defined. The first is to define specific requirements that can be used as guidelines in infrastructure planning and management. The second is to find ways of operationalising the concept of sustainable development in infrastructure performance planning and management. A framework for sustainable urban transport development, based on these needs, will now be given.

From above mentioned definition, the multi-directional conception of sustainable transport development can be stated as:

1. How can basic mobility and accessibility options to people be sustained or enhanced ?
2. How can limited transport related resources, that is environmental, social and economic resources capacities, be used to guarantee intergenerational equity?

Table 2.1 (next page) shows some different levels of the requirements that can be drawn from this multi-directional conception. The level 1 criteria distinguish the two directions in the multi-directional conception, i.e. present development needs and sustainability requirements allow for future development. Level 2 criteria relate, first, to the principal meaning of development, i.e. an improvement in human well-being, economic efficiency and equity (or, an equitable distribution of these improvements over the population) as discussed above. Second, these criteria relate to the use of rate-limited resources, such as emissions per driven passenger-kilometre, which are bounded by a carrying capacity (and thus resilience) of the environmental system, as well as stock-limited resources, such as financial means that can be consumed, but no more than the quantity

⁹This definition is a logical translation of the short definition of sustainable development by the Brundtland Commission on page 32.

Table 2.1: Sustainable development criteria and implications for transport.

Level 1 criteria	Level 2 criteria	Transport-related requirements
Satisfaction of present needs	Improvement in (a) economic and social well-being (b) financial and economic efficiency and (c) equity	Improvement in (a) people and goods mobility, i.e. the ability of people and goods to move or be moved easily and comfortably around; (b) accessibility, i.e. the ability of people and goods to move or be moved easily to essential facilities and services; (c) distribution of mobility and accessibility among people and geographical areas
No hindering of the ability to meet future needs	Sustainable use rate of rate-limited resources Conservation of stock-limited resources	Production rate of wastes is less than what the environment can accommodate Resources consumed are, at any given time, less than the resources available during the time period

available; in other words the financial capacity. Both sets of criteria are accordingly translated into transport-related requirements. The first set contains the development-related requirements as improvements in mobility and accessibility as well as an equitable distribution of these improvements. Followed by sustainability related requirements as wastes production-rate capacities and resource-consumption limits.

Based on the consideration of these requirements, it is postulated that the major sustainable development conditions for transport planning are as follows:

1. Change in well-being with time must not be negative for the social-economic groups in the system under study;
2. Rate of change of each category of the environment, i.e. the rate of addition and/or removal of substances or conditions to/from the environment, must not be higher than its environmental capacity. In this regard, an environmental capacity is the maximum stable rate of change that a specific environment can tolerate to maintain its quality at a desirable level;
3. Rate of consumption of each rate-dependent resource must not be higher than the resource generation capacity, i.e. the maximum long-term rate of generation of the resource that can be reasonably maintained locally;
4. Consumption of each stock-dependent resource must not be more than the maximum allowable amount of resource consumption or the resource consumption capacity.

These definitions have major implications. Two important ones are:

1. Operationalisation of the concept of sustainable development in transport infrastructure planning requires, among other things, the knowledge

and use of (a) locally defined limits to resources consumption and environmental impacts and (b) levels of (transport related) well-being in the community;

2. Different categories of transport system compatibility with sustainable development can and need to be differentiated and defined. To meet this need, three categories of transport systems are proposed. The first category, which is the current focus of many transport professionals and (international) organisations, is a transport system that is sustainable. Based on the above, a *sustainable urban transport system* is postulated to be a transport system that: (a) consumes or requires, at any given time period, rate-dependent resources (e.g. fuel) which are not or will not be greater than the resources that can be internally generated during the time period; (b) consumes stock-dependent resources (e.g. finance) less than the resources consumption capacity; (c) produces, at any given time period, environmental impacts which are not or will not be higher than the environmental capacity, i.e. what the urban environment (human, physical and social-economic conditions) can or will be able to accommodate during the time period. This categorisation can be compared with the single-directional conception of sustainable development as posed in figure 2.3(1), where sustainable development is seen as basically aiming for sustainability instead of sustainable development. The second category is a transport system that is sustainable and developing. A *sustainably developing urban transport system* is postulated to be one that: (a) is sustainable at any given time period; and (b) delivers, developing levels of mobility, accessibility and safety of movement for all social-economic groups at any given time period. A developing level is one that is (a) better than the levels during the preceding time period and (b) less than what the local community considers to be an ideal level. The third and last category is a *sustainable and developed transport system*. This is postulated to be an urban transport system that (a) is sustainable at any given time period; (b) delivers, for all social-economic groups and at any given time period, levels of mobility, accessibility and safety of movement which are equal to what the society considers to be ideal levels.

This categorisation indicates that a sustainable transport system, which is the current focus of many transport professionals and (international) organisations, may not necessarily be the desirable system in many urban areas. The desirable one should be either developing or developed. It is expected that the proposed categorisation of a transport system in terms of whether it is sustainable, sustainably developing or sustainable and developed will serve several purposes. For example, it should help to describe and explain the differences in the current situations in many cities. In addition, it should be useful in the assessment of proposed strategies in among others accessibility planning.

Above discussions can be summarised in a schematic representation of sustain-

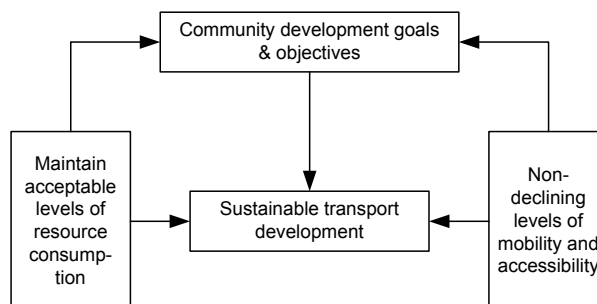


Figure 2.6: A schematic representation of sustainable transport development.

nable transport development (see figure 2.6). Here, communities aim at certain transport development objectives and goals in their transport plans, while maintaining non-declining levels of transport system performance as well as keeping resource use levels below maximally acceptable levels. This figure is based on figure 2.2, but with the resilience effect of the different systems included in the resource capacities of these systems, which are represented by the acceptable levels of resource consumption. Compare also figure 1.9 with the proposed alternative transport planning process.

The proposed conceptualisation of a sustainable and developed transport system reveals the hypotheses that the desirable levels of transport system performance, say mobility and accessibility, in an urban area are significant factors that are compatible with the form, quality of life and environmental integrity in the urban area. In general, a community in an area permanently strives for improved levels of mobility and accessibility over time, ultimately approaching some ideal level in which case all travel related needs of the people are met through the transport system. As communities (hence authorities) become more aware of the environment, a sustainable level of transport system performance in an area implies an improved level of transport system performance that at the same time is not using more resources than available and is not exceeding its environmental capacity. Accordingly, at a given level of environmental capacity the relationship between a sustainable level of transport system performance and available or affordable resources capacity (stock limited items as available finances and land) can be presented by a sigmoidal-curve. The same yields for the relationship between a sustainable level of transport system performance and environmental capacity at a given level of resource capacity. Combining both, this can be graphically depicted as in figure 2.7. Resource capacity might increase over time due to for example economic growth, whereas environmental capacity can change as a result of technology improvements, or through changing political commitment. Both capacities then determine the sustainable level of transport system performance.

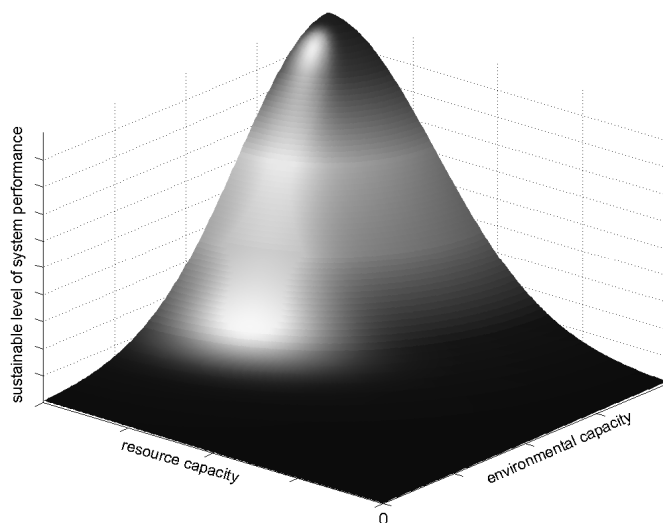


Figure 2.7: Sustainable level of system performance vs. resource capacities.

2.3.2 A different transport planning paradigm

Internalising principles of sustainable development into traditional transport planning theory is complicated by some inherent conceptual differences, which may have become clear from previous discussions. Hence, table 2.2 (next page) provides an overview, based on Zuidgeest and Van Maarseveen (2000), which is not necessarily complete, of some of these complicating aspects, mainly related to dimensions of time and space as well as focus.

Strategic transport planning studies typically have a time horizon of 10 up to 15 years, whereas sustainable development implies intergenerational justice that at least takes up 30 years. Furthermore, transport models mostly give a ‘snapshot’ projection of the situation in an area at a cross-section in time, whereas in sustainable development the process itself is considered important, implying time-dependent models. In addition, in contrast to the use of most transport models, which only have limited exogenous inputs, sustainability studies can always be seen in the light of the larger, global, system. The same applies to hierarchy; transport studies are mostly conducted at a local level and to a lesser extent at a regional, national level etceteras, whereas sustainability studies initially focussed on a world scale, although nowadays more and more regional and also local studies are initiated. Lindquist (1998) justifies using sustainable development at a local transport planning level by stating that most politics is done at the same local level, although the enormity of problems is far beyond the scope of local planning, similar to some arguments in Camagni et al. (1998), as mentioned before. Besides, transport models are typically sectoral models. Travel demand is calculated for a given static, and fixed land-use pattern. Sustainable development models are multi-sectoral and

Table 2.2: Complicating differences between transport planning and sustainable development.

Aspect	Transport planning	Sustainable development
Time scale	10 to 15 years	Intergeneration (>30 years)
Time	Static in time	Dynamic in time (process)
Spatial	Local problems, local solutions	Think global, act local
Hierarchy	Local, regional, national, ...	Global, continental, regional, ...
Disciplinary	Sectoral	Integral (holistic)
Data	Quantitative (model output)	Quantitative and qualitative (indicators)
Approach	Reactive, Predict-provide(-manage)	Proactive (precautionary principle), Predict-prevent, Provide(-manage)-predict

integral, although often limited to only a few sectors. In transport models impacts of alternative plans are calculated and evaluated using performance variables like total vehicle kilometres or total emissions, whereas in sustainability studies, researchers often work with indicators that can be qualitative as well as quantitative. Traditional transport planning theory, often seen as a passive demand-driven system, puts emphasis on meeting the transport demand and therefore maximise utility (see also chapter 1.2), which is also referred to as a *predict - provide - manage*, hence reactive, approach, where through infrastructure provision and/or traffic management predicted travel demand is accommodated. Sustainability studies, however, try to adapt the precautionary principle of *predict - prevent*, or more theoretically try to design (provide or manage) a 'world' that complies to the requirements of sustainability, hence a supply-driven approach, characterised as *provide-predict*.

Despite above mentioned complications in internalising sustainable development in transport planning, the basic sustainable transport development problem is said to be how a transport system:

- can meet non-declining levels of basic movement needs?
- can consume resources equally or less than the available or affordable resources capacity?

From this it is clear that the traditional demand-driven paradigm described in chapter 1 cannot answer these questions and a paradigm shift therefore needs to be advocated that can be characterised as follows (which is a free interpretation from Plaut and Shmueli (2000)):

- Transport planning will be objective-led. A sustainable transport development forms the objective for strategic transport planning;
- Transport planning will be bounded by resources. Resources constitute the bounds that shape transport planning activities;

- Transport planning will be supply-limited rather than demand-driven. The limits to transport growth will be forced upon by engineering and pricing action;
- Transport planning will promote the principles of inter-generational as well as intra-generational equity. Both current and future generations will have to be considered.

Hence, the transport planning paradigm discussed in this thesis is prescribing the future more than it is predicting it. Another, partly similar, shift in thinking and action is proposed by Litman (n.d.a), who claims that sustainability thinking requires rethinking on how transport performance is measured. Transport planners often treat vehicle movement as an end in itself. Sustainable transport planning focuses on access. Sustainable transport planning begins with a community's strategic plan, which individual transport decisions must support. Todd Litman's ideas of what should be in such a plan are summarised in table 2.3 (next page). His focus is clearly on multimodal access and optimal policies.

Summarising, the basic issue, which is considered pertinent for internalisation of the concept of sustainable development, is a validated design of a practical methodology for transport planners and other professionals to determine a set of engineering and/or pricing interventions that contribute to a sustainable transport development. An analytical framework for sustainable urban transport development therefore requires, amongst other things:

1. the determination of movement needs and aspirations of the people and the equivalent desired level of transport system performance;
2. the determination of the resources, which are consumed by transport systems;
3. the determination of available or affordable resource capacities;
4. the design of a suitable tool for assisting the process of developing a transport system meeting adequate levels of mobility and accessibility, of which the resource consumption levels are not higher than the available or affordable resource capacity.

Some specific characteristics of travel and infrastructure related to the first three requirements are discussed in next chapter. The fourth requirement, i.e. developing a modelling tool for sustainable urban transport development that may assist in the planning and design of a transport system that is compatible with sustainable transport development is the topic of chapters 4 and 5.

Table 2.3: Conventional versus sustainable transport planning (Litman, n.d.a).

	Conventional planning	Sustainable planning
Transport	Defines and measures transport primarily in terms of vehicle travel.	Defines and measures transport in terms of access.
Objectives	Maximise road and parking capacity to meet predicted traffic demand.	Uses economic analysis to determine optimal policies and investments.
Facility costs	Considers costs to a specific agency or level of government.	Considers all facility costs, including costs to other levels of government and costs to businesses (such as parking).
User costs	Considers user time, vehicle operating costs, and fares or tolls.	Considers user time, vehicle operating and ownership costs, fares and tolls.
External costs	May consider local air pollution costs.	Considers local and global air pollution, down-stream congestion, uncompensated accident damages, impacts on other road users, and other identified impacts.
Equity	Considers a limited range of equity issues. Addresses equity primarily by subsidising transit.	Considers a wide range of equity issues. Favours transport policies that improve access for non-drivers and disadvantaged populations.
Travel demand	Defines travel demand based on existing user costs.	Defines travel demand as a function, based on various levels of user costs.
Generated traffic	Ignores altogether, or may incorporate limited feedback into modelling.	Takes generated traffic into account in modelling and economic evaluation of alternative policies and investments.
Integration with strategic planning	Considers community land use plans as an input to transport modelling.	Individual transport decisions are selected to support community's strategic vision. Transport decisions are recognised as having land use impacts.
Investment policy	Based on existing funding mechanisms that target money by mode.	Least-cost planning allows resources to be used for the most cost-effective solution.
Pricing	Road and parking facilities are free, or priced for cost recovery.	Road and parking facilities are priced for cost recovery and based on marginal costs to encourage economic efficiency.
Transport Demand Management (TDM)	Uses TDM only where increasing roadway or parking capacity is considered infeasible (i.e., large cities and central business districts).	Implements TDM wherever possible. Capacity expansion only occurs where TDM is not cost effective. Considers a wide range of TDM strategies.

Chapter 3

Sustainable transport development requirements

3.1 Introduction

The paradigm shift in transport planning advocated in previous chapter requires the adoption of a supply-led instead of a demand-driven approach, which is guided by a transport policy objective, as well as sustainability constraints. Hence, sustainable transport development planning is seen as the result of an optimisation process calculating a set of measures to optimise the transport policy objective under system resources constraints, as well as given behavioural rules and mechanisms of the transport system and its users. This process is depicted in figure 3.1.

The travel behavioural rules and mechanism have been discussed in some detail in the transport planning model of chapter 1. The concept of optimisation in transport policy and planning, as well as the model dynamics itself are discussed in chapters 4 and 5, in part II of this thesis. This chapter will discuss sustainable transport development requirements in terms of movement needs and transport system performance objectives. Besides, it will discuss system resources constraints and provides a short overview of possible engineering and pricing measures. One of the shortcomings of the traditional transport planning model, i.e. the lack of sensitivity of travel demand to infrastructure supply, important for the modelling of transport development is also discussed.

3.2 Movement needs and transport system performance

The transport planning process, depicted before in figure 1.4, is often seen as a political process that may include a technical modelling phase, i.e. applying the

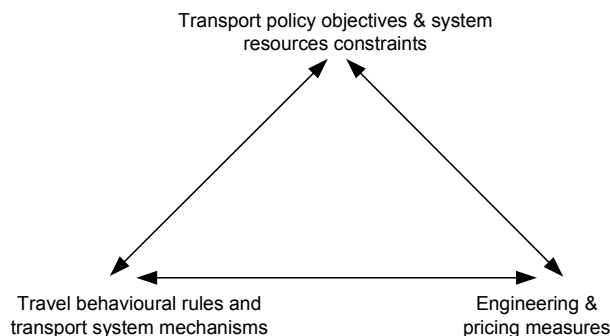


Figure 3.1: Sustainable urban transport planning is objective and resource-led.

transport planning model as described in some detail in paragraph 1.2. A crucial distinction can be made here; the planning process is political, in that interest groups are negotiating solutions in a conflicting arena, whereas the transport model is technocratic, prone to errors and uncertainties, and where the calculated demand - supply equilibrium is unambiguous¹ (Vasconcellos, 2001); the transport model is used as a source of information feeding the political process, i.e. like information on prognoses for future traffic, on current transport problems, as well as on impact of proposed measures etceteras. The transport model is as such not used to answer policy questions in the planning process. In general, transport models have been used to find ways of facilitating future transport needs, as discussed in paragraph 2.3.2, rather than addressing policy directions in planning. In addition, it is the transport modeller, often not the politician, who proposes measures to relieve transport problems. These measures are at best based on the experiences of the persons involved in the modelling and/or political process, but still ad-hoc. Transport modelling techniques that address the political directions in determining the set of measures that need to be taken in order to serve these political directions, given known travel behavioural rules, therefore seem useful. To do so, transport system performance (over time), i.e. the functioning of the transport system at a given level of travel demand and infrastructure supply, should be explicitly related to the political directions or transport policy objective chosen.

This paragraph discusses different approaches to look at transport system performance, and translates them to transport policy objectives. The approaches differ in the way they treat basic movement needs of people, as they range from a mere traffic point of view, that is level-of-service, to social indicators of transport system performance, that is equity. The possible role of transport optimisation models in the political planning process is the topic of next chapter.

¹Compare the overview of critique to the traditional transport planning model that is given in paragraph 1.2.3.

3.2.1 Level-Of-Service

Traditional transport planning tends to focus on accommodating traffic, i.e. the actual movement of vehicles in transport systems. Typical performance related measures are congestion and vehicle speeds. Congestion, that is travel time in excess of that which normally occurs under free-flow travel conditions, is often represented using a link based volume-over-capacity ratio, i.e. V_l/C_l , which is incorporated in the travel time function in equation (1.17). This nonlinear travel time function is the reciprocal of average link vehicle speed s_l , or:

$$s_l = s_l^0 \left[1.0 - \alpha_1 \left(\frac{V_l}{C_l} \right)^{\beta_1} \right], \quad \forall l \quad (3.1)$$

where s_l^0 is the free-flow speed on link l in $[\text{km h}^{-1}]$; V_l and C_l respectively the link volume and link capacity. α_1 and β_1 are parameters [-]. Dimensions of V_l and C_l are in person-car-units (pcu) per hour $[\text{pcuh}^{-1}]$, where all modes are scaled back to equivalent car units².

Closely related to these quantitative measures is the more qualitative measure of level-of-service (LOS) of a link. Here the volume-over-capacity ratio is related to the average speed on a link s_l , which reveals six different levels of quality of traffic movement on a link (A to F). LOS A relates to free-flow conditions ($V_l/C_l \lesssim 0.35$), whereas LOS F relates to heavy traffic ($V_l/C_l \gtrsim 0.90$). In (macroscopic) transport models volume-over-capacity ratios somewhat exceeding 1.0 are generally ‘accepted’, as being representative for heavy congestion that in reality is back-blocking at previous links in the network.

Above mentioned measures relate strongly to motor vehicle traffic conditions. Resources are allocated to improvements that serve these measures. Hence, motor vehicle dependent groups of society tend to receive more resources than those who rely on alternative modes. Link capacity is then increased based on performance measures relating to congestion and speed. Here, it is often (wrongfully) assumed that the traffic volume V_l is independent to changes in link capacity C_l , so ignoring the elasticity of travel demand to changes in transport system performance.

A focus on vehicle travel often results in traffic policies that aim at maximising efficiency of existing infrastructure by maintaining optimal speeds for the highest traffic demand, in other words keeping the LOS at an ‘optimal’ level.

3.2.2 Mobility and person travel

More equitable than measures of speed and congestion are those related to people (and goods). Mobility, in the sense of person travel, is the ability to

²For example, a motor cycle is treated as being 0.5pcu, whereas a bus represents 2.0pcu (for the Dutch situation).

move people, that is mobility improvements that reduce the costs and increase the convenience of moving people. A measure of mobility not only indicates the quality of movement, but also the quantity being moved. Hence, maximisation of mobility requires maximisation of both quality of movement and quantity being moved in a transport network. Therefore, indicators of mobility as for example speed of person movement on a particular route have been proposed for example by the *National Cooperative Highway Research Program* in the USA (NCHRP, 1997), i.e.:

$$\begin{array}{l} \text{Person-speed} \\ \text{[pers km h}^{-1}\text{]} \end{array} = \begin{array}{l} \text{Person-volume} \\ \text{[pers]} \end{array} \cdot \begin{array}{l} \text{Average travel speed} \\ \text{[km h}^{-1}\text{]} \end{array}. \quad (3.2)$$

A corridor mobility index, according to NCHRP (1997), consists then of the speed of person movement divided by some standard value, such as by the person-speed for a link operating at nearly peak efficiency with a ‘typical’ urban vehicle occupancy rate (e.g. 25,000 for an urban street), which is:

$$\text{Corridor mobility index} = \frac{\text{Person-volume} \cdot \text{Average travel speed}}{\text{Optimum facility value}}. \quad (3.3)$$

Related to this is the concept of traveller movement throughput, or person throughput, as is proposed and implemented by Akinyemi (1997) *ibid.* Akinyemi and Zuidgeest (2002). Here, traveller movement throughput for route r , $P_r^{\Delta t}$, in [pers km h⁻¹] is:

$$P_r^{\Delta t} = \Delta t \sum_{m \in \mathcal{M}} (q_{mr}^p \bar{s}_{mr}), \quad \forall r, \quad (3.4)$$

with Δt the time period under consideration (assuming a uniform travel pattern over that time period), \bar{s}_{mr} the average speed in [km h⁻¹] per mode m on a route r and q_{mr}^p the rate at which travellers are being transported per mode along a certain route in [pers h⁻¹]:

$$q_{mr}^p = \theta_{mr} q_{mr}^v, \quad \forall m, r, \quad (3.5)$$

where θ_{mr} is the vehicle occupancy, or average number of travellers per mode m on route r in [pers veh⁻¹], and q_{mr}^v is the flow rate by mode m on route r in [veh h⁻¹]. The average speed \bar{s}_{mr} can also be replaced with \bar{s}_r , indicating an average route speed for all modes m .

The productive capacity of a network is the maximum possible quick and comfortable movement of persons in a network under prevailing conditions. Alternatively, it can be defined as the available network capacity to accommodate quick and comfortable movement of persons by the existing modes of transport during a time period Δt . It expresses the maximum hourly rate at which people

can be conveyed across a given network under the prevailing traffic composition and occupancy conditions. Using equation (3.4) the productive capacity of a route in a period of time Δt is C_r^P in [perskmh⁻¹] is expressed as:

$$C_r^P = \max_{u \in \mathcal{U}} (P_r^{\Delta t}), \quad \forall r, \quad (3.6)$$

with $u \in \mathcal{U}$, being a set of engineering or pricing measures that can be deployed on a route, e.g. introduction of a bus service. Similarly by summation over all routes r this leads to the productive capacity of the transport network C^P , i.e.:

$$C^P = \max_{u \in \mathcal{U}} \sum_{r \in \mathcal{R}} (P_r^{\Delta t}). \quad (3.7)$$

Hence, the productive capacity of a transport network consists of:

1. the mode specific passenger traffic capacity for different routes in the network;
2. the average speed of a mode on different routes in the network.

A speed or congestion based optimal condition on a route or in a network, as in paragraph 3.2.1, is conceptually different from a throughput-based optimal condition of a transport network³. Therefore, traffic policies that focus on mobility or person travel and productive capacity of a transport network aim at reaching desirable or optimal levels of person mobility in a network.

3.2.3 Accessibility

Accessibility is a concept used in a number of fields such as transport planning, urban planning and transport geography. Accessibility can be interpreted as the means or ease by which an individual or group can accomplish some economic or social activity. Hence, accessibility is not necessarily related to a transport system. Another interpretation, however, is that accessibility is the potential for interaction or a measure of the intensity level of the interactions, see for example (Tagore and Sikdar, 1995). A third type of interpretation relates accessibility to network access, i.e. the ease of reaching and/or leaving a transport network or service. The most popular type of interpretation, however, is to express accessibility as the amount of money, time and trouble it costs a person or group to cover the distance from their place of origin to their destination. This diversity of interpretations has resulted in the use of many different measures or indicators. Morris et al. (1979) had earlier categorised them into process and outcome indicators and described the advantages and

³Akinyemi and Zuidgeest (2002) discuss some experimental results that confirm this statement.

disadvantages of each category of measure. Process indicators are based on the supply characteristics of the system and/or individual and outcome indicators are based on actual use and level of satisfaction. The greatest problem (at least for engineers) with most of these interpretations and measures is that they do not adequately or at least explicitly accommodate engineering characteristics such as capacity and level-of-service supplied by the transport system. Hence, emphasis should be given to a process type measure that also takes the supply side of the transport system into consideration for use in engineering science, since it is first of all the ability to move (for all groups in society), not the degree to which this ability is exercised that characterises a mobile individual as well as a mobile society.

In general four different components of accessibility can be identified, as discussed in greater detail in Geurs and Ritsema van Eck (2001):

1. A *transport component*, reflecting the disutility that individuals or groups experience in bridging the distance from their origin to destination using a specific transport mode;
2. A *land-use component*, reflecting the magnitude, quality and character of activities found at each destination and this component's distribution in space;
3. A *temporal component*, reflecting the availability of opportunities at different times of the day and the times at which individuals participate in certain activities;
4. An *individual component*, reflecting the needs, abilities and opportunities of individuals.

From a strategic modelling point of view the first two components are particularly essential for an accessibility measure in that they reflect the characteristics of travel demand and infrastructure supply in an area and hence the expected demand-supply equilibrium. The temporal aspect is most important in disaggregate and operational studies and will not be considered in this research. An individual component on the other hand can at an aggregate level discriminate between different groups of people in society and is therefore important in reflecting equity issues. Hence, a measure for accessibility that both reflects the distribution and importance of activities within the urban area as well as describes the ease of traversing a distance via the transport system for different social-economic groups in society is looked for. From this the early concept of potential of opportunities for interaction suggested by Hansen (1959), who sees accessibility as the 'potential of opportunities for physical interaction between spatially separated activities by a transport system' appears useful at first sight, i.e.:

$$A_i = \sum_{j \in \mathcal{J}} a_{ij} = \sum_{j \in \mathcal{J}} X_j f(c_{ij}), \quad \forall i, \quad (3.8)$$

where A_i is the integral accessibility for zone i in an area and a_{ij} is what Morris et al. (1979) call the relative accessibility, since it describes the relation or degree of connection between two zones i and j . X_j is the attractiveness of opportunities in zone j and $f(c_{ij})$ an impedance function that is monotonously decreasing with increasing travel impedance, for example as the exponential formulation in equation (1.7). Some authors, for example Wachs and Kumagai (1973), have extended above model by distinguishing different modalities, hence obtaining the mode m specific accessibility A_{im} for zone i .

Walter Hansen's model can be criticised in that it only represents one out of two attractiveness involved in the interaction between zones, as is claimed by Sales Filho (1998). Laerte Sales Filho therefore replaces X_j in equation (3.8) with a geometric average of the origin and destination attractiveness:

$$A_{ij} = (\tilde{Q}_i \tilde{X}_j)^{\frac{1}{2}} f(c_{ij}), \quad \forall i, j. \quad (3.9)$$

In this formulation origins and destinations are coupled, hence indicating the relative importance of this specific origin-destination relation. The values of \tilde{Q}_i , and \tilde{X}_j are corrected values for the production potential Q_i for zone i and attraction potential X_j for zone j respectively, in order to obtain a more realistic level of interaction. This correction is done by taking the average number of possible interactions. In addition, the extent to which an origin or destination contributes to accessibility also depends on its relative attractiveness compared to the total sum of origin or destination attractiveness. Taking these considerations in mind the original values for origin and destination attractiveness are corrected as follows:

$$\tilde{Q}_i = Q_i \frac{\sum_{i' \in \mathcal{I}} Q_{i'} + \sum_{j \in \mathcal{J}} X_j}{2 \sum_{i' \in \mathcal{I}} Q_{i'}}, \quad \forall i, \quad (3.10)$$

as well as:

$$\tilde{X}_j = X_j \frac{\sum_{i \in \mathcal{I}} Q_i + \sum_{j' \in \mathcal{J}} X_{j'}}{2 \sum_{j' \in \mathcal{J}} X_{j'}}, \quad \forall j. \quad (3.11)$$

The geometric average consideration in this way not only takes account of the dimensional problem but also shows the combined effect of both attractiveness in the composition of the 'potential for interaction' (Sales Filho, 1998).

The spatial interaction model in equations (3.9) to (3.11) is illustrated in figure 3.2 (next page). Here, the boxes indicate the production potential for zone i and attraction potential for zone j , separated by a travel impedance $f(c_{ij})$. Zones i and j are surrounded by zones x and y with their respective production potential Q_x, Q_y and attraction potential X_x, X_y .

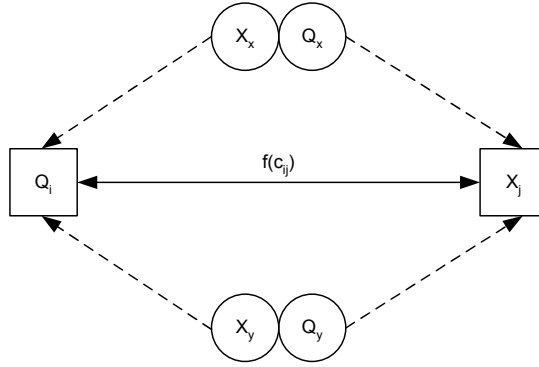


Figure 3.2: A spatial interaction model with travel impedance $f(c_{ij})$.

The generalised cost c_{ij} depicted in equations (3.8) and (3.9) can be (and are) replaced with a so-called composite cost, or logsum cost c_{ij}^* , which is discussed in paragraph 3.3.1.

Equation (3.9), also incorporating the corrections on Q and X , can be made multimodal m and user specific, by applying a population segmentation k , as follows:

$$A_{ijm|k} = (\tilde{Q}_{i|k} \tilde{X}_{j|k})^{\frac{1}{2}} f(c_{ijm|k}^*), \quad \forall i, j, m. \quad (3.12)$$

The integral accessibility as in equation (3.8) for zone i can then be calculated using the following summation:

$$A_i = \sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}} \beta_{2m} \sum_{k \in \mathcal{K}} (\tilde{Q}_{i|k} \tilde{X}_{j|k})^{\frac{1}{2}} f(c_{ijm|k}^*), \quad \forall i. \quad (3.13)$$

A higher A_i now implies an improved accessibility for zone i . Parameter β_{2m} is added as a weight for mode m , through which a political or a planner's preference can be expressed. A total systems accessibility is logically formed by obtaining a supertotal:

$$A = \sum_{i \in \mathcal{I}} A_i. \quad (3.14)$$

Using equations (3.9) to (3.11) a so-called accessibility matrix A_{ij} can be constructed, which is, though somehow related, conceptually different from the origin-destination table, T_{ij} , as depicted by equation (1.5). The latter concentrates on trip volumes, which is the way in which the interactions are distributed, whereas the accessibility matrix deals with inter-zonal potentials of opportunities for interaction, that reflect transport - land-use combination effects and difficulties involved through travel impedance.

The capacity of an individual or group of people to meet their basic needs is dependent largely on his/her or their accessibility to various economic, cultural and social activities. The above mentioned accessibility measure, based on Sales Filho (1998), can now be used to analyse transport system performance from an accessibility point of view. Transport policies that focus on accessibility of a transport network aim at maximising interaction opportunities of people.

3.2.4 Equity

As expressed before in this research sustainability is not only about environmental amenity due to transport improvements in an urban area; it must also be about increasing human potential and well-being for *all* groups in society. Transport improvements should reduce social exclusion, hence increase equity. Four types of transport social exclusion are identified in DETR (2002):

1. *spatial exclusion*, when people simply cannot get to the location they wish to access;
2. *financial exclusion*, when they cannot afford to get there;
3. *temporal exclusion*, when they cannot get there at the appropriate time;
4. *personal exclusion*, when they lack the mental or physical equipment to handle the available means of mobility.

From a strategic modelling point of view, the first two components are in particular essential. The extent of these components can be influenced through aggregate supply-side interventions, as capacity and cost improvements to all social-economic groups in the area. The third and fourth type of social exclusion typically involve disaggregate and operational analysis of transport supply. These components relate to what Litman and Burwell (n.d.) call vertical equity, which implies that accessibility options should improve for people who are economically, socially and physically disadvantaged. This in contrast to horizontal equity that refers to the distribution of negative externalities as pollution as well as costs and benefits amongst causers and non-causers, i.e. get what you pay for and pay for what you get. These concepts can be further elaborated by introducing equity of opportunities and outcome, spatial equity, social equity as well as economic equity. A good overview is provided by Geurs and Ritsema van Eck (2001).

One of the few examples of integrating equity principles in transport modelling can be found in Meng and Yang (2002), who treat equity as a constraint that requires the travel cost ratio between before and after a road improvement for any origin-destination combination to be less than a certain threshold value ψ . They, however, don't make any distinction in population segments, but instead focus merely on spatial equity for the different origin-destination pairs. A slightly adapted version of this constraint is:

$$\max_{i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{K}} \left\{ \frac{c_{ij|k}^*}{\bar{c}_{ij|k}^*} \right\} \leq \psi, \quad (3.15)$$

with $\bar{c}_{ij|k}^*$ the composite cost of travel on origin-destination pair (i, j) for group k , before implementation of some transport measures, and $c_{ij|k}^*$ the composite costs afterwards, see paragraph 3.3.1. The threshold is: $\psi < 1.0$, if all users in the network benefit from the transport project considered. Consequently, if: $\psi > 1.0$, then there is at least one origin-destination pair and/or group of users who will experience travel costs greater than before, hence inequity of the transport project is revealed. If: $\psi = 1.0$, then nobody experiences benefits or loss due to the transport project. An effective interval: $[\min_{k \in \mathcal{K}} (\psi_{\min}^k), \max_{k \in \mathcal{K}} (\psi_{\max}^k)]$, can also be considered to distinguish the groups (with segmentation dependent threshold ψ^k) that experience highest benefit, versus those that experience lowest benefit, as well as the magnitude of this difference. Hence, the benefit distribution associated with a certain project can be studied.

The approach followed in equation (3.15) doesn't distinguish who is actually benefiting from an improvement; it only prevents users to be worse off than before. Hence, a weak version of horizontal equity.

As a measure of equity \mathcal{E}^Q , constraint (3.15) can be reformulated as a measure reflecting the mean deviation between composite costs in the network by the different population segmentations, as follows:

$$\mathcal{E}^Q = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \left(\frac{1}{K} \sum_{k \in \mathcal{K}} \left| \left(\frac{Q_{i|k}}{\sum_{k' \in \mathcal{K}} Q_{i|k'}} \right) c_{ij|k}^* - \bar{c}_{ij}^* \right| \right), \quad (3.16)$$

with \bar{c}_{ij}^* being the weighted average composite costs of travel on origin-destination pair (i, j) for all population segmentations k , which is:

$$\bar{c}_{ij}^* = \frac{1}{K} \sum_{k \in \mathcal{K}} \left(\left(\frac{Q_{i|k}}{\sum_{k' \in \mathcal{K}} Q_{i|k'}} \right) c_{ij|k}^* \right), \quad \forall i, j. \quad (3.17)$$

A lower value of this measure guarantees a more equitable or levelled distribution of transport opportunities to the different segmentations. The weighted average composite costs acting as a source-term, could however also imply that investments are only directed at levelling the absolute deviation towards zero only. An overall improvement of the transport system somehow is not reflected in this measure; only the distribution of transport opportunities.

Transport policies that combine the accessibility formulation in equation (3.12) with the equity constraint or equity measure aim at the prevention of both

strategic types of social exclusion, through group-specific accessibility enhancements, with the weak horizontal equity idea of fairness of cost and benefit allocation between the user groups.

3.3 Transport and development

Many factors contribute to economic and social development, but mobility and accessibility are especially important because the ingredients of a satisfactory life, from food and health to education and employment, are generally available only if there is adequate means of moving people, goods and ideas (Owen, 1987).

To suggest that transport leads to development is to adopt a supply-led approach and there is a school of thought, which argues that at early stages in the development process, and by implication in many developing countries, it is the provision of transport, which leads to a widening of markets, increased production and associated multiplier effects of an economic and social nature (White and Senior, 1983). An alternative view to this supply-led approach is that transport provision is invariably a response to demand and it is rarely developed except where there is a demand. This view directly serves the type of transport demand that is called *revealed demand* as expressed in the trips that are actually made; compare figure 1.5. However, at any place and at any point in time there is likely to be an element of *latent demand* that is the existing demand which cannot be satisfied, perhaps because of inadequacies in the infrastructure or prohibitive cost, and which might be called delayed demand, but also completely new demand that may be created by additional or improved infrastructure (Hilling, 1996). Latent demand is very likely to be significant in developing cities. Although, the question may rise whether some element of latent demand can be the reason for the transport system to be developed, it is assumed in this research that there is a positive relation between transport system performance and social-economic development. That there might also be a permissive relationship that is transport does not in itself stimulate economic growth, but is such that it does not inhibit such growth when other stimuli are operating, or even a negative relationship, that is when returns on investment in transport are less than from the same investment directly into productive activity, is not considered here, but is discussed in Gauthier (1970).

The capacity of an individual or group of people to meet their basic needs is dependent largely on his/her or their accessibility to various economic, cultural and social activities, as discussed in paragraph 3.2.3. Hence, in response to an improvement in accessibility from infrastructure investment (e.g. road capacity increase, rail expansion etceteras), Banister and Berechman (2000) claim that firms and households can increase their demand for infrastructure facilities (i.e. the trip generation effect). They may also change their trip pattern (i.e. the trip distribution effect); or their choice of travel mode and their travel route (i.e. the modal choice and traffic assignment effect). They may relocate (i.e. the spatial location effect), or they may adopt all of these options. In turn, each

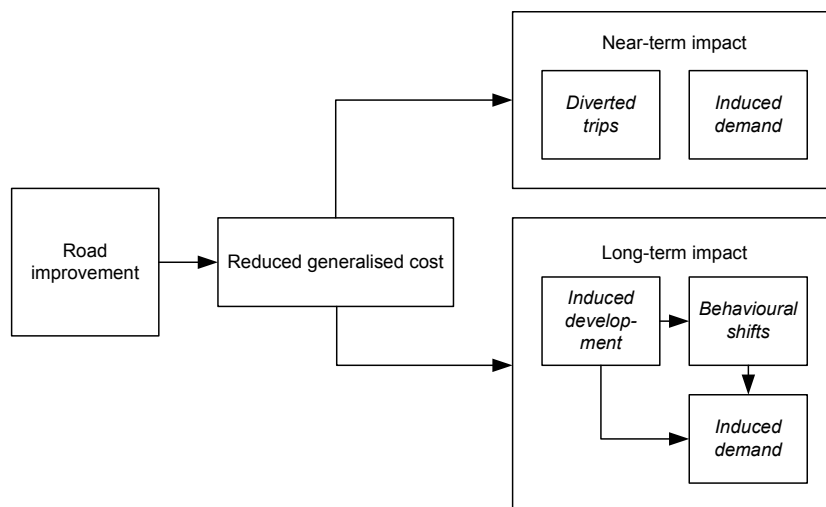


Figure 3.3: Road improvement and induced travel demand, adapted from Cervero (2002).

of these effects will influence the degree of use (hence the performance level) of the transport system, which is in existence, or which is being constructed or expanded. Likewise, it is hard to show what newly created traffic on a road is due to a trip generation, trip distribution or modal split effect, which is *induced demand* or *generated demand* (new trips, mode shifts and longer trips), or what traffic is caused due to the traffic assignment effect, which is *diverted traffic* (route shifts and temporal shifts). This is what Cervero (2002) calls near-term effects of road improvements. At the longer-term the spatial location effect may stimulate *induced development* (land-use shifts), hence *behavioural shifts* (new levels of vehicle ownership and transit usage) and again induced demand (new trips, mode shifts and longer trips). These near-term and long-term improvements are depicted in figure 3.3, based on Cervero (2002). Here a causal relationship is shown, where road investment increases travel speeds and reduces travel times, which influences the generalised cost of travel, as in equation (1.2). Hence, induced and diverted traffic can be expected in the near-term. At the longer-term spatial location effects may lead to more induced traffic. These causal relations can be expressed in terms of the elasticity of the demand function, as seen in figure 1.5⁴. Goodwin (1996), for example, found the following elasticities for vehicle travel with respect to travel time, for urban roads, -0.27 in the short-term and -0.57 over the long-term. SACTRA (1994), in addition, reports for the same type of elasticities -0.50 respectively -1.0 . Hence, a 10% decrease in travel time, would lead to a 5% increase in traffic in the short-term.

⁴Then elasticity is expressed as the quotient of the relative change in an effect parameter and the relative change in a suspected or actual causal parameter.

Generally, only the negative aspects of induced demand are considered. Evidence on whether building new roads, or expanding existing ones, creates new traffic has been the topic of a considerable and passionate debate. Especially since the appearance of the UK government-commissioned *SACTRA report* (SACTRA, 1994), many articles were published claiming evidence for demand inducing effects of new transport capacity, i.e. Goodwin (1992), Goodwin (1996), Prakash et al. (2001), Noland and Lem (2002) as well as Cervero (2002) and including estimated time-lags in Noland (2001). To what degree and under what circumstances latent traffic will be induced, is, however, still a matter of debate.

In view of many cities in developing countries, but also in developed countries, it is also worth seeing the benefits of induced trips. People do not make trips if the generalised or composite cost, a disutility, of making those trips is greater than the associated utility of performing the activity (see equation (1.14)). Therefore if these costs were reduced, some additional induced traffic would turn up, since accessibility has improved, hence more social-economic opportunities can be realised. Benefits of accommodating latent demand are discussed in Dahlgren (2003). Significant benefits arise from reductions in schedule delay⁵, delay on alternate routes, because of route shifts and also because of new trips, although these latter trips constitute a relative small number, at least in developed countries.

From above discussion it should be clear that the equilibrium framework due to Manheim is 'alive and well'. Infrastructure supply and travel demand are in a rather complex equilibration process. Reduced generalised costs result in new or diverted traffic, which is increasing generalised cost in return, hence suppressing demand again. This vicious circle of demand inducement should also be taken in mind when developing transport planning methodologies.

3.3.1 Dynamics in travel demand

Travel demand is derived and facilitates a complex and spatially varied set of activities as work, shopping, recreation and home life, but is also regarded to be elastic to changes in the transport system as seen in previous discussion.

To make the actual (revealed) travel demand responsive or elastic to changes in travel cost for specific groups, *potential travel demand* Q_i is defined as the travel demand that exists on the basis of the information on land-use and social-economy characteristics for zone i in the study area.

The revealed travel demand $T_{i|k}$ is then a function of for example the household characteristics for population segmentation k in zone i , i.e. $H_{i|k}$, and a measure of accessibility by group type, $A_{ijm|k}$, as in equation (3.12), thus revealing:

⁵That is leaving earlier than desired in order to reach one's destination in time, perhaps arriving earlier than desired.

$$T_{i|k} = f(H_{i|k}, A_{ijm|k}), \quad \forall i, \quad (3.18)$$

which is analogue to equation (1.2). $T_{i|k}$ is assumed to be monotonously decreasing with deteriorating accessibility. By doing so, trip rates are influenced by the level-of-service in the network.

Potential travel demand Q_i is said to be produced in the origin-zones i of the area under study. The capacity of an origin zone to produce (or generate) these trips is considered to be dependent on land-use and social-economic characteristics, for example through the home-based variables population, vehicle ownership and area of residential land, as well as for example non-home based variables as retail employment and area of retail land.

Trip attractions for the destination zones j are constrained to equal total trip productions. Thus, it is only the relative proportions of attractions that are of significance⁶. Hence, their real values are used. Indicators such as zonal employment and school places are considered to be good indicators for trip attractions. If considered important, a categorisation is possible into several population segmentations k , for example stratified for income, therefore obtaining potential zonal travel demand per segmentation k , i.e. $Q_{i|k}$.

Estimating potential travel demand poses some challenges. A standard revealed preference (RP) survey might reveal enough information, even though it only obtains data on the actual trip making. A clustering study of equal social-economic classes, though experiencing different accessibility, hence actual trip generation, might reveal such information. However, problems encountered when trying such an approach using an extensive household survey for Metro Manila, The Philippines, have recently been reported in Bierman (2004). Another option is to use locational models at the land-use level as discussed in BTE (1998), where the trip generation number is compared with the actual trip making, given the actual level of accessibility. The difference reveals information on the latent demand. Further research on the estimation of potential travel demand will be proposed in chapter 7.

Based on above discussion it can be said that the potential travel demand is assumed to consist of a part revealed demand and a part latent demand. Latent demand increases if accessibility in the transport system decreases and vice versa, as could be seen in equation (3.18). The potential demand $Q_{i|k}$ for zone i , which equals $f(H_{i|k})$ in equation (3.18), is monotonously decaying using a nonlinear decay value $D_{ij|k}$, after De la Barra (1989), similar to the demand functions presented in Bell and Iida (1999):

$$D_{ij|k} = a_{1|k} + b_{1|k} \exp(-\lambda_{2|k} c_{ij|k}^*), \quad \forall i, j, \quad (3.19)$$

⁶Here, it is assumed, as is generally done, that trip productions are more easily and accurately estimated than trip attractions, see equation (1.4).

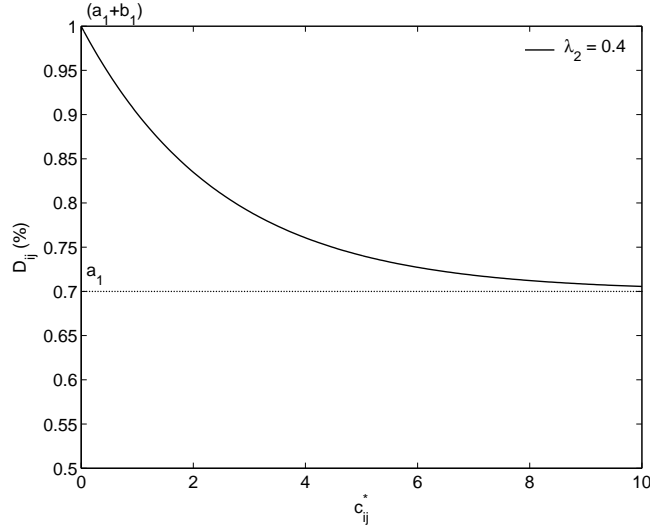


Figure 3.4: Elasticity in the generation of trips, based on De la Barra (1989).

where $c_{ij|k}^*$ are the composite costs of travel, $a_{1|k}$ being the minimum ratio of trips (to $Q_{i|k}$) which will be performed (so-called *captive trips*, irrespective of composite costs) and: $a_{1|k} + b_{2|k} = 100\%$, the maximum ratio of the trips that can be performed at ideal circumstances, or free-flow conditions. The parameter $\lambda_{2|k}$ indicates the sensitivity of the amount of revealed travel demand between two zones to the composite costs. Figure 3.4 shows this elasticity in the generation of trips, assuming 70% captive trips.

The composite cost, or logsum cost, on an origin-destination pair (i, j) is calculated by aggregating over the choice-alternatives n (for example modes and/or routes) that serve the origin-destination pair. Hence, reflecting a kind of level-of-service, or *average perceived benefit*, over all modes and routes from to the trip origin to the destination. The composite cost function $c_{ij|k}^*$ can be derived, from the discrete-choice model (1.12) with alternatives for both mode m as well as route choice r :

$$p_{(m,r)|ij:k} = \frac{\exp(-\lambda_{3|k} c_{ijmr|k})}{\sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} \exp(-\lambda_{3|k} c_{ijm'r'|k})}, \quad \forall m, r, \quad (3.20)$$

taking the integral over $c_{ijmr|k}$:

$$c_{ij|k}^* = \int \frac{\exp(-\lambda_{3|k} \check{c}_{ijmr|k})}{\sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} \exp(-\lambda_{3|k} \check{c}_{ijm'r'|k})} d\check{c}_{ijmr|k}, \quad \forall i, j, \quad (3.21)$$

yields⁷:

$$c_{ij|k}^* = -\frac{1}{\lambda_{3|k}} \ln \left[\sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \exp(-\lambda_{3|k} c_{ijmr|k}) \right], \quad \forall i, j. \quad (3.22)$$

Parameter $\lambda_{3|k}$ represents the sensitivity of the composite cost to the different utility values of the individual choices, where a high sensitivity is for example indicated with a value of: $\lambda_{3|k} = 0.6$ compared to a low sensitivity if: $\lambda_{3|k} = 0.2$. More generally, the composite cost function represents the expected maximum utility or minimum disutility of a given choice set, where $\lambda_{3|k}$ is the scale parameter, as in equation (1.12). The trip maker ranks the options in terms of perceived costs, and will choose the one that minimises such costs. This interpretation has the advantage of being closer to classical demand theory, in which consumers relate quantities of goods or services with prices.

Closely related to this concept is the notion of a *consumer surplus* indicator (seen as a measure of welfare). Changes in consumer surplus resulting from changes between two different policies, say policies (1) and (2), are compared, as the difference in benefit, or benefit ratio ΔS , i.e.:

$$\Delta S = -\frac{1}{\lambda_{3|k}} \ln \frac{\left[\sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \exp(-\lambda_{3|k} c_{ijmr|k}) \right] (2)}{\left[\sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \exp(-\lambda_{3|k} c_{ijmr|k}) \right] (1)}. \quad (3.23)$$

Trip makers can now for example travel at a lower price (disutility) and thus enjoy an increase in the consumer surplus. To calculate the total benefit, ΔS should be multiplied by the total number of (new) trip makers, who are actually benefiting from the increase in consumer surplus.

Several authors have proposed different formulations to the elastic demand problem, somehow trying to model the transport system dynamics of figure 1.5; for example in an early paper by Friesz and Fernandez (1979), a good overview in Sheffi (1985) or recent advances in Bell and Iida (1999).

Ortúzar and Willumsen (2001), in addition, provide a discussion on the problems encountered when trying to model (in particular the calibration and

⁷De la Barra (n.d.), based on (Williams, 1977), mentions three conditions for the use of the composite cost concept that is:

1. $c_{ij|k}^* < \min_{(m,r)} c_{ijmr|k}$, which states that the composite cost must be less than the measurable generalised cost of the best option;
2. $\lim_{\lambda_{3|k} \rightarrow -\infty} (c_{ij|k}^*) = \min_{(m,r)} c_{ijmr|k}$, which implies that, as the absolute value increases, the composite cost tends towards the cost of the best option;
3. $\lim_{\lambda_{3|k} \rightarrow 0} (c_{ij|k}^*) = 0$, which sets the lower bound for the value of the composite cost by stating that such value cannot be negative.

Checks on these conditions are built-in in the model presented in chapter 5.

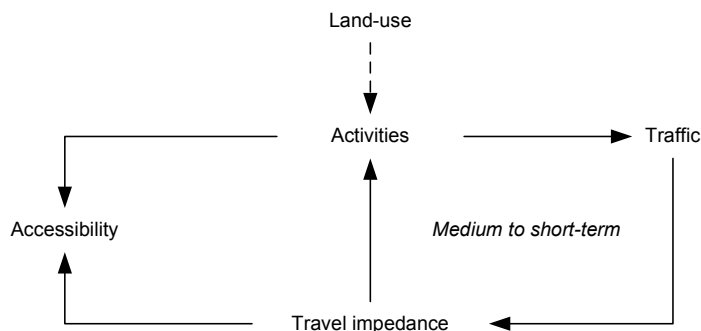


Figure 3.5: Equilibration of travel demand and infrastructure supply vs. accessibility.

validation issues) these effects of network changes on trip generation. The two greatest problems with such model dynamics are:

1. Model dynamics itself. The time lagged adjustment of people to changes in the transport system is not modelled in traditional static travel demand models. The so-called stimulus-response relation is not well understood (see paragraph 5.2.2). Ortúzar et al. (2000) discuss for example how Stated Preference (SP) techniques can be used to obtain a better understanding of this particular aspect of model dynamics;
2. Modelling with longitudinal instead of cross-sectional data. With the use of cross-sectional data the dynamics of transport systems can hardly be captured. For example the disequilibrium state of the transport system and the lagged adjustment are not revealed using cross-sectional data. Collecting longitudinal data, that is repeated cross-sectional data and/or panel data, however, suffers from high costs, non-response, drop out, panel conditioning etceteras. A thorough discussion on the use of longitudinal data is provided in Goodwin et al. (1990).

The best of these different methods for revealing data on model dynamics are still a matter of debate. In this research no attempt is made to choose a best method.

Modelling the equilibrium framework depicted before in figures 1.3 and 1.5 creates the opportunity of studying the effects of transport improvements on the transport patterns and travel rates in the network. It is shown how, from exogenous land-use derived, activities relate to travel demand, hence to traffic and travel impedance. The amount of revealed demand is based on the travel impedance (or composite cost) in the transport network. An equilibrium level of traffic results. Transport system performance in terms of accessibility is related to the potential number of activities related to the subsequent composite costs of travel linking activities, as depicted in figure 3.5. The state of equilibrium that occurs is of medium to short duration. As a consequence of transport

improvement, new demand for travel may occur, which in turn, generates a subsequent need for new infrastructure facilities. An important aspect of the equilibration of travel demand and infrastructure supply is that this equilibration doesn't occur instantaneous and, in general, requires considerable periods of time to transpire. The main reasons for this are the long periods necessary for investment implementation as well as the time needed for the demand side adjustment. A disequilibrium framework for modelling this dynamic process based on the issues discussed in this and previous paragraphs is implemented in chapter 5.

3.4 Resource consumption and capacities

As discussed extensively in chapter 2, sustainable urban transport development requires that resources constitute the bounds that shape the development. Hence, the development paths of the transport system are limited by resource capacities as environmental and financial capacities. Transport professionals have several types of measures at their disposal to control or affect these *development* or *transition paths*, which are discussed in this paragraph.

The most common form of pollution regulation is through setting environmental standards. These standards imply the establishment of particular levels of environmental concentration for a certain pollutant. They are normally set with reference to some health-related criterion. This criterion is related to the environmental capacity of the system, which can be defined as the capacity of ecosystems to absorb human activity. Air quality standards are set by governments or international organisations as the Intergovernmental Panel on Climatic Change (IPCC) and WHO, mostly in terms of maximum concentrations for pollutants. For an overview on European concentration standards the reader is referred to Whitelegg (1993), or more general to Schwela and Zali (1999).

Targets for carbon dioxide (CO₂) emissions, the main greenhouse gas, are often stated as reductions in the cumulative share of transport in CO₂ emissions. An example being the *Kyoto protocol* as mentioned in chapter 2.3. Environmental capacity objectives can obviously be formulated for one, possibly dominant, proxy for a pollutant, or a (weighted) combination of several pollutant types. It is also possible to define specific targets for emission reductions for transport systems in demarcated geographical areas in terms of the *local environmental transport capacity*, which is the maximum level of transport, or maximum transport capacity in order to consider a transport network acceptable in terms of the environmental objectives. This objective is then reciprocal to the environmental capacity objective.

As an example for modelling emission of pollutants $p \in \mathcal{P}$, Nagurney (2000) developed transport models using mobile or *non-point source emission* estimates, rather than stationary sources, which is the relationship that the volume

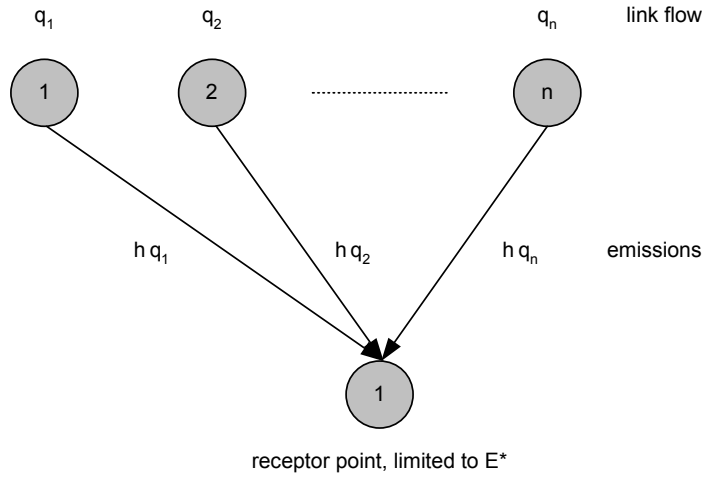


Figure 3.6: Spatial dispersion of vehicular emissions, based on Nagurney (2000).

of emissions is equal to the product of a composite emission factor h^* (allowing for different types of emission factors) multiplied by the link vehicle flow q_l^v . By doing so there is a single receptor point associated to the emissions: $h^* q_l^v$. This concept of spatial dispersion of vehicular emissions is illustrated in figure 3.6. The total local environmental capacity, or desired environmental standard is indicated as E^* . If a distinction is made in mode specific composite emission factors h_m , a possible environmental goal may now be:

$$\sum_{l \in \mathcal{L}} \sum_{m \in \mathcal{M}} h_m^* q_{lm}^v \leq E^*. \quad (3.24)$$

A single receptor point is regarded reasonable as the scale of application, the urban transport network, is rather limited. The environmental standard is ideally also set at this urban scale for reasons of urban pollution control. If necessary multiple receptor points could be defined, each related to a bundle of link flows.

Likewise, fuel consumption models can be used to put a constraint on the total fuel used by the traffic in the transport network. Roughly it can be said that fuel consumption per unit distance is high at low speeds and decreases as speeds increase. Fuel use is also known to be proportional to emissions of carbon monoxide (CO), volatile organic compound (VOC) as well as nitrogen oxides (NO_x) emissions. Both are monotonously-decreasing functions with travel speed (roughly up to about 80 km h^{-1})⁸. Nitrogen oxide emissions (NO_x), however, increase again from travel speeds in the range of 50 km h^{-1} . These

⁸Often a distinction is made in speed ranges. At higher speeds emissions tend to increase again.

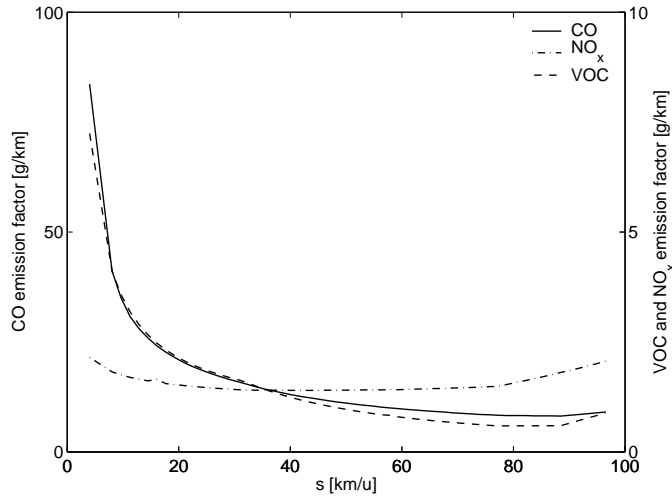


Figure 3.7: Variation of emission factors with average speed, after Al-Deek et al. (1995).

relations are depicted for CO, VOC and NO_x gasoline emissions for a representative mix of vehicles in figure 3.7, based on Al-Deek et al. (1995), who estimated pollutant emission factors by average speed (up to about 80 km h^{-1}) from the emission factor model MOBILE5A. Similar patterns exist for CO_2 emissions and fuel consumption. Both can be approximated by monotonously decreasing functions with average vehicle speed. Hence, the emission factors h become functions of the average link speed s as well.

Applying general least-squares techniques (see appendix B.1) the following power functions for CO as well as VOC are estimated as:

$$h^{\text{CO}} = 240.5s^{-0.82}, \quad (3.25)$$

$$h^{\text{VOC}} = 16.5s^{-0.68}. \quad (3.26)$$

For NO_x emissions the following quadratic polynomial can be fitted:

$$h^{\text{NO}_x} = 0.0003s^2 - 0.0272s + 2.0103. \quad (3.27)$$

All three curve-fits show a goodness-of-fit ratio (see also appendix B.1): $R^2 > 0.9$, which implies a good representation of the data-points. However, it should be noted that the data-points themselves, obtained from Al-Deek et al. (1995), are derived using a software program that by itself is based on curve-fits as well, instead of from actual end-of-pipe emission measurements. Furthermore,

it should be noted that these data-points were derived by simulating route-choice behaviour and environmental effects by introduction of advanced traveller information systems (ATIS) focusing on the year 2003. In other words, circumstances in terms of temperature, vehicle mix and age, control measure implementation etceteras comply with the research objectives of that research, not necessarily this one. However, it is believed that these equations still represent the right order of magnitude, useful for the type of strategic modelling intended in this research.

Another example for dealing with environmental capacity can be found in the environmental capacity oriented urban traffic strategic planning method as proposed by Huapu and Peng (2001), who distinguish a total environmental capacity per exhausted gas. Apart from this they also put inspiration out of the influential concept of *ecological footprint* developed by Wackernagel and Rees (1995). Ecological footprint is then a sustainability index based on land use area. It quantifies the intensity of resource consuming related with regional bearing capability. In their research, Huapu and Peng (2001), define the capacity of ecological footprint of an urban passenger transport system as the environmental capacity, analogue to the use of E^* in equation (3.24). Likewise, in a recent paper by Shrestha et al. (2005), a modal split that meets NO_x targets for the city of Beijing, China is determined. Determining the ‘available’ environmental capacity on an demarcated urban scale, however, poses some challenges, as is discussed in Van Nes (2004), who recently studied scale-problems in defining and deriving environmental capacity.

An earlier related example is discussed in Immers and Oosterbaan (1991), who developed a model for the calculation of environment-friendly traffic flows in urban networks on the basis of a constrained user-equilibrium assignment. The constraints are formulated as a maximum traffic environmental capacity (per road link) for noise pollution as well as air pollution as compared to standard road capacity.

In addition, Akinyemi (1998) distinguishes, besides an environmental capacity, also a financial capacity, i.e. unit costs of construction, maintenance and operation of infrastructure, e.g. c^U or link specific unit costs c_l^U , multiplied by the quantity of construction, maintenance and operation at the link levels, U_l , which should not exceed the allocated budget, F^* , as well as a social capacity constraint, interpreted as a measure of traffic unsafety per passenger kilometre. These capacities can be equally treated as the mobile emission concept of Anna Nagurney in equation (3.24), for example for the financial capacity:

$$\sum_{l \in \mathcal{L}} c^U U_l \leq F^*. \quad (3.28)$$

3.4.1 Measures for controlling transition paths

Several transport policy measures can be identified that are used for transport planning activities, often combined in an integrated approach, in which case a whole range of interventions are applied simultaneously. In an extensive study, described in May et al. (2001), some 60 transport policy measures have been categorised in five groups, being:

1. Land use measures;
2. Infrastructure provision (for car, public transport, non-motorised transport and/or freight specific);
3. Management of infrastructure (for car, public transport, non-motorised transport and/or freight specific);
4. Information provision (for car, public transport, non-motorised transport and/or freight specific);
5. Pricing (for car and/or public transport specific).

Through these measures, transport system performance (LOS, mobility, person travel, accessibility and equity) can be regulated as well as the environmental, financial and social externalities that result.

Furthermore, a distinction can be made between *demand-side* oriented policies and *supply-side* oriented policies. For the traditional transport model, in particular, a few possible regulators, or controls in managing the equilibration process of travel demand and infrastructure supply can be distinguished, based on the categorisation (between brackets) in table 3.1.

Demand-side options are generally influencing the generalised cost of travelling through out-of-pocket costs (affordability), hence controlling total motor vehicle travel demand, by amongst others (TRB, 1997):

- Road use fees and parking tax;
- Vehicle ownership and acquisition taxes;
- Inducements for ride-sharing and tele-commuting.

Supply-side options deal with availability of transport options and capacity in the transport network, by amongst others (TRB, 1997):

- Investment in public transport and non-motorised transport modes;
- Highway capacity and traffic flow improvements;
- Intelligent Transport Systems (ITS) technology.

Table 3.1: Instruments for regulating transport system performance.

Demand-side	Supply-side
Affordability of transport (5) - fixed travel cost - variable travel cost	Availability of transport options (3,4) - mode options - route options - mode prioritisation - traffic management Road capacity (2) - construction - maintenance - mode prioritization

The (effectively) available road capacity to certain modes can be regulated through road construction, pavement maintenance as well as mode prioritisation measures. This effective road capacity then influences the generalised cost of travelling through the travel time. In addition, the number of travel options available in combination with the travel time per option influences the composite cost of travel. Hence, for example, accessibility can be controlled.

Part II

Modelling sustainable urban transport development

Chapter 4

Optimisation as a tool for sustainable transport policy design

4.1 Transport policy and planning

The interaction between travel demand and infrastructure supply is highly complex, since both demand and supply comprise of and are influenced by different and complex as well as interacting processes, as seen in chapter 1. Hence, the change in a single transport policy measure can be difficult to understand and predict. Furthermore, the decision-making, leading to these policy measures is part of another complex process, the policy process.

In this chapter optimisation is discussed as a tool for designing optimal transport policies in view of these complex and interacting processes. Optimisation will (conceptually) prove to be able to assist in the reasoning from a policy objective to the policy measures taken.

4.1.1 The transport policy process

Transport policy making according to Tolley and Turton (1995) can be defined as ‘the process of regulating and controlling the provision of transport to facilitate the efficient operation of the economic, social and political life of the country at the lowest social cost.’. This implies, amongst others, providing adequate and sufficient infrastructure capacity and efficient operations to best meet the current and future needs of the stakeholders involved (those are: inhabitants, companies, travellers, transporters and/or policy makers). Transport is therefore regulated and controlled through public policy. The level of regulation and control, however, may vary from full-state interventions to free-market

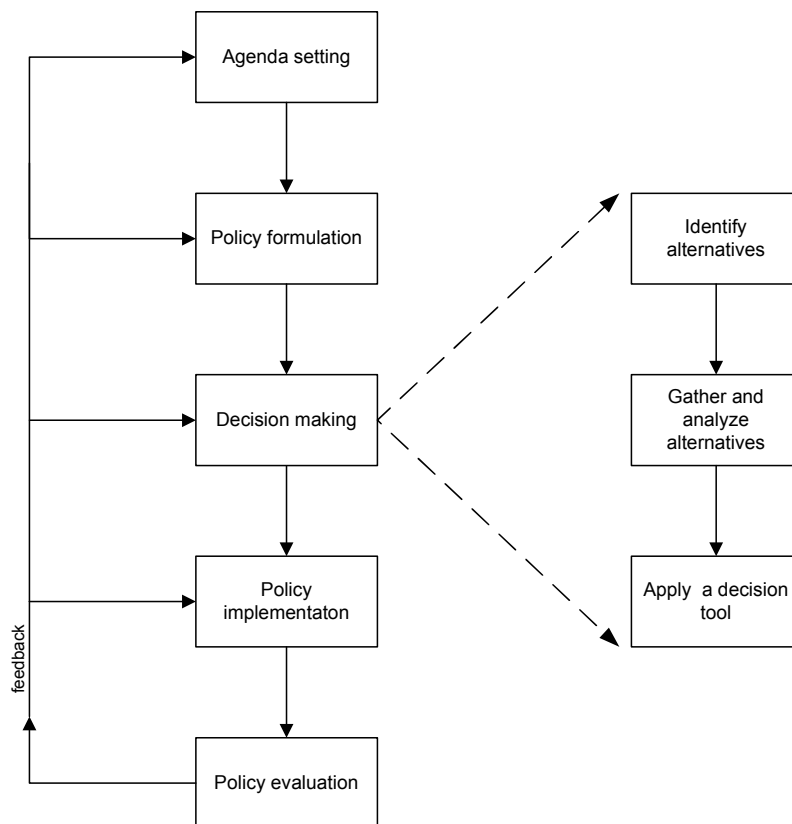


Figure 4.1: The public policy cycle and formulation/decision-making stage.

principles¹.

Public policy, which encompasses transport policy, is often depicted as a top-down iterative process, typically including: (1) agenda setting, (2) formulation, (3) decision-making, (4) implementation, and (5) evaluation stages, that move in a linear fashion from higher to lower levels of government. This concept of political sciences is many times accredited to Lasswell (1951), and depicted schematically in figure 4.1. The first step, agenda setting, is the process by which issues come to the attention of policy makers. The second step, policy formulation, is the process in which options are formulated, risks, costs and benefits are determined, technical evaluation is performed and a ranking of options is made. In the third step, which in some literature is integrated with the policy formulation step, the decision is made whether one, none or

¹The role of free market versus public interest through the state is not discussed in this thesis, although it will be clear that the conceptualisation of sustainable transport and subsequent modelling in this thesis are assuming some kind of regulatory powers, most probably through government.

some combination of solutions is selected based on a reasoning from policy objective(s) to solution(s) or measure(s). This step encompasses the sequential decision-making process where: (1) alternatives are identified, (2) alternatives are gathered and analysed, and (3) finally selected with the aid of a decision tool, as can be seen in figure 4.1 also. It is normally assumed that in this decision-making process, technical experts as transport planners come in to participate in the decision-making (Barkenbus, 1998). In the fourth step of policy implementation, policies are put into effect. Lastly, the chosen policy is monitored and evaluated by the authorities themselves as well as the public in the fifth step of policy evaluation. Final judgement on the success of policies can then be made.

As said before, much work in transport research is in some way meant to advice policy makers taking the third decision-making step². A role and magnitude of involvement, which is highly dependent on the complexity of transport policy objectives as well as the adopted decision-making approach, which are discussed in the next paragraph.

4.1.2 Decision-making approaches

Decision-makers are faced with a decision that must be made or a problem that must be resolved. This is at the heart of policy analysis, which is to evaluate, order and structure incomplete knowledge so as to allow decisions to be made with as complete an understanding as possible of the current state of knowledge, its limitations and its implications (Morgan and Henrion, 1990). As there is not such a thing as a single model of decision-making that can be used to guide a planning process in transport, Meyer and Miller (2001), as do Ortúzar and Willumsen (2001), argue that the role and character of the technical transport planning process (basically encompassing the second sequential decision-making process in figure 4.1) within transport policy design highly depends on the decision-making approach followed. Furthermore, the type, context and complexity of the transport problem, as well as the organisational and political context are dominant in the type of decision-making. In general, three sets of approaches can be distinguished (summarised from Meyer and Miller, 2001; Ortúzar and Willumsen, 2001; Anderson, 2003):

Rational actor or normative decision-making approach Here a rational and completely informed group of decision-makers is assumed. In this well-structured and data-intensive approach, the decision is based on maximising the attainment of a set of goals and objectives, sometimes applying mathematical programming techniques. Compare also figure 4.1;

Satisficing or behavioural decision approach Here it is assumed that decision-makers aren't completely informed nor are they utility maximisers

²Barkenbus (1998) even argues that transport experts are too much focused on this step in the policy cycle.

as in the normative approach, but rather satisficers searching for acceptable, though rational, but not necessarily optimal, solutions. Mostly, only marginal improvements to the status-quo are made, though still based on some selection of goals or objectives. Hence, also called incrementalism;

Group decision approaches This group of approaches sees decision-making as a group learning process with several stakeholders with (different levels of) decision authority, but diverse goals, values and interests, hence reflecting organisational or political settings. Negotiation, bargaining and compromising between groups occurs, through which technically non-optimal solutions often result. Qualitative, quantitative information and forecasts are used to support the overall decision-making process.

Transport models may be applied in all three approaches, in an attempt to provide analytical insight in the decision-making process. The actual use differs between approaches, mainly reflecting the amount of rationality allowed for in the transport planning paradigm, hence picturing the transport planner from an objective, analytical expert (technocrat) in the normative approach to a communicative expert with technical knowledge skill (mediator) in the group decision approaches³. Transport models can assist in proposing the measures complying with the stated goals and objectives. However, transport planners (involved in the technical decision-making stage) do not have a tool for deriving measures from an objective readily available. Most transport models can only evaluate proposed measures and not derive them from objectives. So, transport planners themselves will have to come up with an (initial) set of measures that (might) meet the objectives, before a model can be used to evaluate. This can be a problem if policy objectives become more and more complex; As they do.

4.1.3 Complex transport policy objectives

Most of the goals and objectives⁴ in transport policy and planning seem very hard to realise. For example, one can look at the (concept) mobility plan for Flanders, Belgium (Ministerie van de Vlaamse Gemeenschap, 2001) that explicitly mentions as one of its objectives: ‘The essence of a sustainable mobility development is to facilitate the interaction-needs by people and freight that are a resultant of social and economic activities, through a transport system in such a way that the demand for mobility can be fulfilled now and in the future.’. Three constraints that go with this objective are accordingly posed, assuming:

1. efficient use of scarce resources;

³An interesting discussion on transport planning paradigms and the amount and type of rationality allowed for, is discussed chronologically in Talvitie (1997) and Willson (2001); the latter one including a comment by Antti Talvitie (Talvitie, 2001).

⁴The words goals and objectives are used interchangeably here, although one should realise there are conceptual differences between a goal and an objective. A goal should be seen as the end-result aimed for, whereas an objective is a (quantifiable and) identifiable step towards a goal.

2. equity to all groups of users;
3. obedience of capacity limits of external effects, like safety, liveability as well as environment.

Plans such as the Flanders mobility plan usually mention a set of traffic and transport related measures to realise their objectives. However, the proposed measures in these plans (in the end) often appear not sufficient to fully meet the objectives. Transport planning rarely optimises around one objective, usually there are multiple (sometimes contradictory) objectives or even non-viable objectives.

Goodwin (1999) illustrates this by showing that objectives in transport planning have become increasingly complex, while discussing the whole policy development from predict-provide approach, to sustainable transport for the British transport policy. He shows that during the decades of predict-provide, the formulation of a suitable policy was easy. The classical four-step transport planning model gave a prediction of the growth and the suitable policy was to build enough roads to provide for it. Every transport planner could formulate such a policy. But for a sustainable transport system it is less clear what measures have to be taken, when, where and by how much.

Dependent on the decision-making approach chosen, policy analysis can be applied to obtain better insight and understanding on problems with complex objectives in several ways. Three motivations for this are taken from Morgan and Henrion (1990). First, assuming (transport) policy analysis is all about having clearly defined and authorised decision-makers, who have well-defined goals or objectives as in the normative approach, policy analysis is employed to determine how best to achieve these complex objectives. Secondly, in some cases the alternative policy options are clearly established and the point of analysis is to assist the decision maker in choosing among a discrete or continuous set of options to find the 'best' one from the alternatives, as in the satisficing approach. Thirdly, the possible alternatives are not clearly defined and analysis is used to help the decision maker to identify and explore possible alternatives as well as to choose among them, useful in group decision approaches.

In transport policy analysis, decision makers are also confronted with an increasing number of clearly or less clearly defined, interacting as well as complex transport policy objectives and planning restrictions. A transport planning method that can assist in determining sets of transport policy measures, which give maximum value to these transport policy objectives, hence designing a transport planning, will prove to be useful.

4.2 Transport planning as a design problem

An interesting definition of planning in general is provided by Intriligator and Sheshinski (1986) who refer to it as 'the elaboration of an explicit set of de-

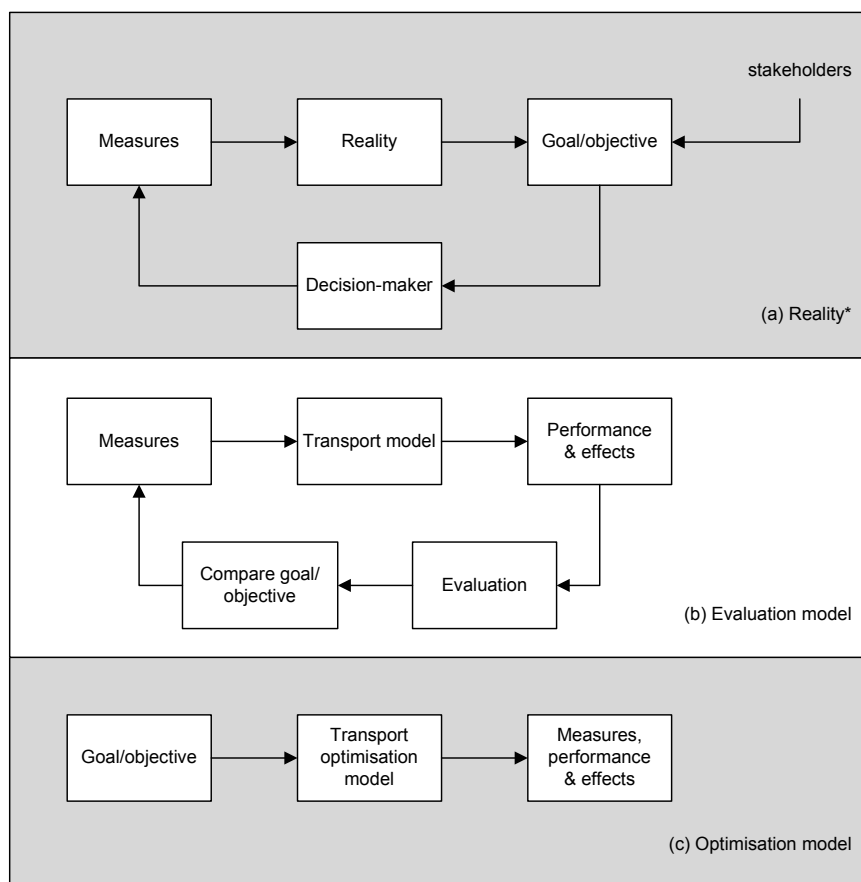
cisions concerning the present and future values of certain choice variables by a decision maker (planner) in order to achieve certain goals'. Planning according to them, therefore, involves the determination of a strategy that, in turn, involves decisions on both actions and their timing or pattern of implementation. Relating this to transport planning, which in paragraph 1.2.1 has been characterised as both steering the process of allocating traffic generators over space and time as well as the provision of transport facilities, it is clear that an important part of (transport) planning is the reasoning from explicit goals and objectives down to actions over time and space, given some behavioural rules.

Transport planning is as such instrumental to transport policy. Stakeholders, as policy makers, define goals and objectives. The decision maker weights these and compares the weighted result with observed reality. Based on this comparison, decisions on how to act are made. Hence, the whole process is built on personal assumptions of how policy measures affect reality (Knoflacher et al., 2000), as illustrated in figure 4.2(a). Apparently, some optimisation criterion (goals/objectives like 'lowest social cost' or 'best') as well as possible measures or interventions ('infrastructure capacity' or 'efficient operation' that may be individually applied or integrated), are inherent to the design of transport policy and planning.

Engineering design is an iterative process, where new design proposals are generated and evaluated, with the iterative part consisting of *synthesis*, *simulation*, *evaluation* and finally *decision* (Roozenburg and Eekels, 1998). Each design can be evaluated using a simulation model and then compared to some pre-defined goal/objective. If the design does not meet the requirements it is modified and evaluated again in search for the best possible design. Hence, design is essentially optimisation, since optimisation can be defined as 'the procedure or procedures used to make a system or design as effective or functional as possible, especially the mathematical techniques involved' (English dictionary, 2000).

In transport planning a lot of (technical) knowledge on transport systems is available. This knowledge has been formalised in traditional transport models with their known predict-provide-manage characteristics. For use in transport planning, these models are meant to evaluate a-priori, as can be seen in figure 4.2(b) as well. The idea of an optimisation model is to use the knowledge about transport systems not just to evaluate alternatives, but to design alternatives on the basis of some pre-defined goal or objective, as seen in figure 4.2(c).

Mathematically, a policy objective can be seen as a cost criterion that needs to be maximised or minimised. Measures that are to be taken, can be seen as the input of a certain dynamical system $\mathcal{S}(\mathcal{T}, \mathcal{W}, \mathcal{B})$. In this system, \mathcal{T} is the relevant time axis, \mathcal{W} is the signal space and \mathcal{B} the behaviour. \mathcal{B} is the set of trajectories that are compatible with the phenomena modelled by the system. The signal space, or manifest variables \mathcal{W} can be divided in two parts; first, an input part that can be freely chosen and, second, an output part that is



* adapted from Knoflachner et al. (2000).

Figure 4.2: (a) Reality*, (b) evaluation modelling versus (c) optimisation in transport planning.

completely determined by a given input.

In terms of a transport system \mathcal{S} , \mathcal{W} is a factor containing variables like the number of trips and the travel time. \mathcal{B} is a collection of all the possible combinations of \mathcal{W} , mostly represented by equations. For the transport system, this can be a modal split function (as in equation (1.12)) or a cost-flow function (as in equation (1.17)). The input of the system contains the measures that can be taken to influence the transport system.

The optimal solution is then the input, at which the cost criterion has the highest value. This way, an optimal package of measures is found to comply best with a certain policy objective, hence a transport planning is designed. Doing so, it is also possible that the model confirms that the system aimed for

is unrealistic, infeasible or ill conceived. If the optimisation doesn't converge to a desired system, the concept has to be modified or the problem and alternatives reformulated, possibly also resulting in new possible measures, adjusted constraints or even a different objective function.

Hence, employing control and optimisation techniques to transport planning design, can in principle lead to improvements compared to current transport planning practice. However, there will always be parts in the design and decision-making process that require human or qualitative judgement, no optimisation technique can do. The definition of the model equations, objective function as well as constraint values need to be done before the actual optimisation can take place. This definition phase is prone to errors. At best the optimisation model outcomes can guide here in the reformulation of equations and constraints, in other words the input part of the system.

Chapter 5 discusses a dynamic optimisation model in detail. But first some more content will be given to the pro's and con's of using optimisation techniques in transport planning design.

4.2.1 Existing optimisation models

Optimisation techniques have been applied extensively in transport science, although mainly for the purpose of traffic management studies (i.e. advanced control techniques for optimal adaptive traffic control), transit studies (i.e. optimal line - and routeplanning in traffic systems) and in logistics (i.e. the travelling salesman problem). Optimising for the design of tactical or strategic transport plans is done less frequently, although a great body of literature exists on traffic assignment techniques (see for an overview on static assignment techniques amongst others (Ortúzar and Willumsen, 2001), on dynamic assignment techniques (Friesz and Bernstein, 2001), and on transit-assignment (De Cea and Fernandez, 2001)) as well as transport network design (see amongst others Steenbrink, 1974; Sheffi, 1985; Bell and Allsop, 1998; Yang and Bell, 1998; Bell and Iida, 1999; Nagurney, 2000), which will be discussed in paragraph 4.3.3. Optimisation is also common in the field of transport economics. Few of these optimisation models will be mentioned here, as they are illustrative for the development of general transport optimisation models.

In *The Economics of Welfare* (c.f. (Pigou, 1932) in Crozet and Marlot (2001)), the economist Arthur Pigou introduced the idea of a congestion toll by coming up with the famous two-road example. This can be seen as the first transport optimisation model. This approach has been studied extensively and many different models have been made in connection to this idea. Optimal road pricing measures are the output of these models, and can be seen as the optimal input for the controlled transport system. An example of such a study can be found in Hau (1992). Timothy Hau presents a model a model that studies the theoretical benefits of road pricing based on first economic principles. As Hau's method is only graphical and is not based on known parameter values, it

is perhaps less suitable to a policy maker who wants to formulate an alternative for a particular transport system.

Indeed most models in transport economics have the problem that they are less suitable for modelling a real world system. A step in the right direction is the model of Yang and Huang (1998), based on the same idea as Arthur Pigou's, but also applicable to a transport network. They present a model for optimal road pricing for a transport network with some bottlenecks. This is an important feature for a designer of alternatives. Yet, the model has only one objective (welfare maximisation) and one type of measure (road pricing).

Models that can handle more objectives and measures, while still being applicable to real world problems as well, are seldom. An example is the model of De Borger and Wouters (1998). They made a model using static optimisation techniques. In relation to transport policy, this model has some useful characteristics. It is meant for a whole country or region, which makes it more suitable for transport planning. It covers both public and private transport and it calculates optimal prices and capacity for both modes. The model is limited in two directions. First, it is not a dynamic model, which means that a planning process during some years can hardly be modelled. Second, the network and causal relations are not modelled explicitly. This means that interpreting the results for implementation in the real world is not very easy.

The discussion of these three models illustrates what is being looked for, i.e. a model that is easy to interpret, has a direct link to real world transport networks, and can cope with different measures and objectives. In chapter 5 these characteristics will be derived systematically. It is not strange that existing models do not have these characteristics. Most of them are meant for researching some policy objective or for finding a theoretical best situation. This is something different from a tool for designing alternatives. Some more discussion on the implementation of the concept of optimisation in transport planning will be given in next paragraph.

4.2.2 Transport planning as an optimisation problem

There is, however, some critique to the previously mentioned concept of optimisation in transport planning. According to Willson (2001) for example, transport planning cannot be characterised as a rational reasoning from an objective to a measure as in optimisation. Richard Willson thus characterises the decision-making process in transport planning as a group decision approach that involves negotiation, bargaining and compromising, different from merely optimisation. Because of all the political interests involved in the whole process of transport planning, Willson claims it is better characterised as a communicative process (Willson, 2001). Therefore, this seems contradictory with the rational or normative decision-making approach characterised by optimisation, system dynamics and the linear fashioned policy cycle.

Also from an engineering design point of view, optimisation can be criticised. Roozenburg and Eekels (1998) studied the design process of objects. For them, designing is a creative process with no single best solution or guaranteed reasoning resulting in a good design. They show that it is theoretically impossible to deduct a certain form directly from a certain function. And this deducting is exactly what happens when you are designing through optimisation. On the contrary, Andersson (2001), claims (referring to Roozenburg and Eekels (1998)) engineering design is essentially an optimisation problem as long as some human or unquantifiable judgement is allowed for at some stage.

There are many transport experts that criticise the predictive value of transport planning models as well. Willson (2001), once more, argues that these prediction problems make transport models less suitable for transport planning.

Although most of the criticism is right, it is possible to make an optimisation model with such properties that it can positively handle most of the comments. Nevertheless, there will always be certain limitations. As long as these limitations are respected, the model can well be used in transport planning. Some of the critiques will be discussed in terms of whether they are relevant for the problem at stake or how they can be avoided.

The first criticism by Willson (2001) is about the policy cycle. The policy cycle is based on a rational and powerful policy maker who can dictate the whole policy process. The rational and powerful policy maker, however, is a seldom case in transport policy. In reality, policy makers have to negotiate with stakeholders and even work together with them as in a group decision process. But even in that case optimisation can be useful. This can be illustrated with two cases. First, a leading policy maker with certain objectives surrounded by some reactive, but powerful, stakeholders. The policy maker will try to formulate a policy considering the other stakeholders or formulate a policy co-operating with the other stakeholders. Second, a complex policy surrounding where a variable number of stakeholders will co-operate and negotiate, resulting in a certain policy.

In the first case, the policy maker, seen as the leading stakeholder in this process, will have to come up with a concrete alternative, matching his or her transport-related objectives. For this, optimisation can be used, in order to be sure that the alternative is, given the policy objective, the optimal solution. However, this does not mean that this alternative will be implemented, since objectives and constraints coming from other stakeholders with different interests will have to be considered as well. Looking at the policy cycle this means that the phases following the policy formulation are not the exclusive domain of the policy maker. The policy maker's alternative will have to be evaluated with respect to all stakeholders' interests and, depending on the results, probably be adjusted accordingly. In practice this process may be somewhat fuzzier than described, but the main point is that the existence of other stakeholders does

not decrease the necessity for a policy maker to formulate alternatives to fulfil objectives, and optimisation is a good way to do this. Also it makes it easier for the policy maker to argue his or her alternative and make the causalities between measures and objectives clear and explainable to the other stakeholders. This is especially important when he wants to co-operate with the other stakeholders rather than considering them as opponents.

In the other case with a group process involving more stakeholders, things are not really different. All the stakeholders will formulate alternatives according to their (most probably complex) policy objectives, possibly considering the reaction of other stakeholders as well. The result of the process will depend on politics more than on transport causalities. Nevertheless, all stakeholders will have to find at least one alternative that fully or partly meets their objective. In the end, a package of measures gets implemented, partly of fully meeting the stakeholders objectives. So, despite the fact that both the alternatives proposed by a stakeholder and the alternative implemented after the process can differ from the alternative dictated by the optimisation model, this model is still helpful in formulating alternatives complying with the objectives as well as providing an important input to the negotiation process. In some cases multi-objective optimisation can even be an useful option (see for an interesting discussion on this topic Andersson (2001)).

The second criticism, by Roozenburg and Eekels (1998), is on the nature of the design process. They consider the design progress as a creative process. A creative process cannot go together with optimisation, they reason. But there are some differences between the design problems of Roozenburg and Eekels (1998) and the design of a transport planning. The difference between designing, for instance, a water cooker and designing a transport planning is the number of possible shapes or forms. The design of a chemical plant with one reactor may serve as an example as well. Two raw materials are added resulting in an output of one product. Temperature and input as well as output of the raw materials are variable. The design problem is to find a temperature with a certain input, the shape or form, resulting in a certain quality and quantity output, its function. Still, there are infinite possible designs, but the situation is different from designing a water cooker, where the degrees of freedom are far outreaching those for operating the chemical plant. It is the same with designing a transport policy. The degrees of freedom are limited by some fundamental points of departure in terms of infrastructure, modes, impedances, land-uses and time. Hence, it must theoretically be possible to model close-to-all possible alternatives⁵ and choose the best one. As can be done by optimisation.

Taking the reactor and the design problem again, it is clear that this is not the design problem according to Roozenburg and Eekels (1998). There is not a lot

⁵Particularly, when considering strategic modelling, where the alternatives are mostly specified in terms of general variables as capacities or prices.

of creativity in the reasoning from function to form. The specification of the design problem results in one optimal alternative.

So, two ways of designing can be distinguished, one with a creative component resulting in a satisfying, not-necessarily best, design, and one with optimisation resulting in a best design. Related to this, if no suitable alternative is found in the optimisation process, the designer can be sure there is none; the design problem formulation might be unrealistic, infeasible or ill-conceived. The only option is then to reformulate the design problem. In the case of Roozenburg and Eekels (1998) it is possible that a functioning form does exist, but the designer cannot find it. Another difference is the phase of the modelling in the design process. In the case of optimisation, modelling starts before alternatives are formulated. In the case of evaluating, first the design is made, before the modelling starts (compare again figure 4.2). Related to this is the fact that in the case of optimisation all possible measures have to be described mathematically. The key question therefore is whether designing a transport policy alternative is more alike Roozenburg and Eekels (1998)'s design problem or more like optimisation.

At first sight transport planning design seems to have more in common with Roozenburg and Eekels (1998)'s design problem. Measures as ramp metering, increase of comfort in public transport, educational programs for drivers, etc. are almost impossible to model as components of a system. Also, the mechanisms by which these measures influence the policy objective are difficult or even impossible to describe in a mathematical model. But also the design model of Roozenburg and Eekels (1998) has a phase in which the design has to be modelled, i.e. the evaluation phase. This phase is also present in transport policy design. After designing the alternative, it is evaluated *a-priori* in a transport model. And if the set of measures can be modelled after designing the alternative it should be possible to do this before. The evaluation of alternatives by modelling is merely done by translating the proposed measures into more generalised or strategic measures that can be implemented in a (strategic) transport model. The measure on ramp metering for example is translated into the measure road capacity increase. If that can be done, it should also be possible to model the generalised measures in an optimisation model and translate them into specific measures after the optimisation process. This should work well because the aim of the optimisation process is to design an alternative and not to create a blueprint. Hence, it appears to be sufficient to know a certain policy direction with system information on for example capacity increases, road pricing, public transport capacity and frequencies and visions for spatial planning. These kinds of measures can be described in a mathematical model, even in continuous-time, hence containing all possible alternatives.

The third criticism points at the predictive value of transport models in general. Why optimise when there are serious doubts about the rightness of the dynamics? This problem can be dealt with by looking at other reasons for modelling, apart from predicting. Modelling helps to write down all knowledge

and assumptions about transport systems systematically, so that it is internally consistent. For the optimisation model, this results in a system that is consistent with the way a policy maker looks at the system. The model does not claim a perfect prediction. However, it claims that if the assumptions of the policy maker are true, the measures will have the predicted results. As long as the policy maker realises that the model is only a result of his or her own assumptions, an optimisation model can be used.

The above mentioned comments on the general criticisms on optimisation in (strategic) transport planning have direct consequences for the type and characteristics of the optimisation model. These comments are repeated here, but on a somewhat more technical level that is relevant for choosing the form of the mathematical model in chapter 5:

Integrated model: The aim of the model is to help with complex causal reasoning from objective to set of measures. The idea is that different processes, interacting and even contradicting with each other, influence these causalities. This means that all relevant traffic and transport related processes and their interactions have to be modelled. The only realistic way to do this is to make an integrated model;

Suitable for policy making for real world problems: It is not realistic to ask for models to include all relevant processes for detailed transport networks subscribing measures on a similar detailed scale. Nevertheless, the whole exercise is meant to result in an intervention in the real world. This means that the translation from the optimisation result to a real world intervention must be done without losing validity, perhaps applying hierarchical modelling techniques as will be mentioned in chapter 6. In practice, this will mean that typical characteristics of real world networks will have to be modelled, but unnecessary details can be avoided. For example, if road pricing is introduced, traffic will relocate to toll-free roads. This process has to be modelled. But this can be done without modelling a detailed network by making one hypothetical alternative road if the decision maker is merely interested in quantifying expected toll revenues. The result is that the effect is clear and plays a role in finding the optimal solution, but the model stays simple and understandable:

Clear and explainable causalities: In the case of more stakeholders negotiating and co-operating to reach a certain transport policy, the stakeholders will need to explain their preferred measures and the logic behind their ideas. This means that the optimisation model not only has to produce a suitable alternative, but also an explanation why that particular alternative is optimally related to a certain objective. A simple reference to an optimisation model will of course not be enough. A black box model will not be suitable. The causalities should be clear and should be described without difficult mathematical theory. This does not mean that the process of policy making is considered completely rational. It only means

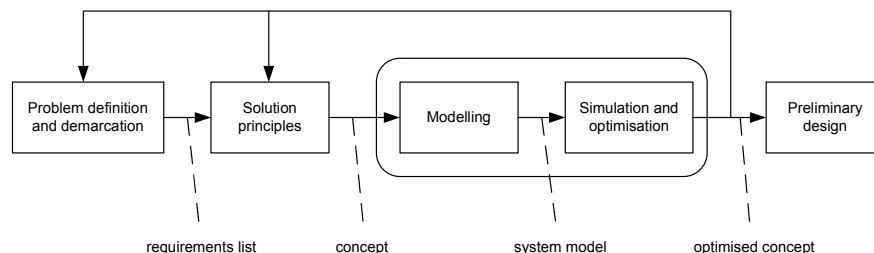


Figure 4.3: The system design process, based on Andersson (2001).

that any stakeholder will connect measures and objectives by causalities and want to explain these to other stakeholders. So, a model that helps to design alternatives should also be capable of giving the corresponding explanation.

The system design process that results is summarised in figure 4.3. In the problem definition and demarcation phase the design problem and its limitations are defined and on the basis of this, a list of requirements is made that will be used to generate solution principles, (strategic) design alternatives, or concepts, which will guide the modelling and optimisation phase. In this phase, the design problem is formulated at a sufficient degree of refinement. Simulation and optimisation are then deployed to predict the properties of the system design. The output is called a preliminary design in order to show these are suggestions that still can be altered through human and unquantifiable judgement, or more detailed at a tactical or operational level of modelling. The output can also be the reason to reformulate the problem and solution principles.

Referring back to this chapter as well as to chapters 2 and 3, the sustainable urban transport development problem can be defined as a (constrained) optimisation problem. Next paragraph will discuss modelling and optimisation approaches for sustainable transport that have been found in literature, before proposing a different model in next chapter.

4.3 Methods and models for modelling sustainable transport

A classification of existing methods and modelling approaches to sustainable transport development is discussed in this paragraph. Several approaches can be found in literature, all having a different perspective on or interpretation of sustainability. Hence, sustainability is weakened here to include some goal or objective for sustainability, which is therefore fundamentally different from the standard forecasting approach in transport planning, as discussed before in paragraph 1.2. The approaches differ in whether they include dynamics in time and whether development trajectories are discerned.

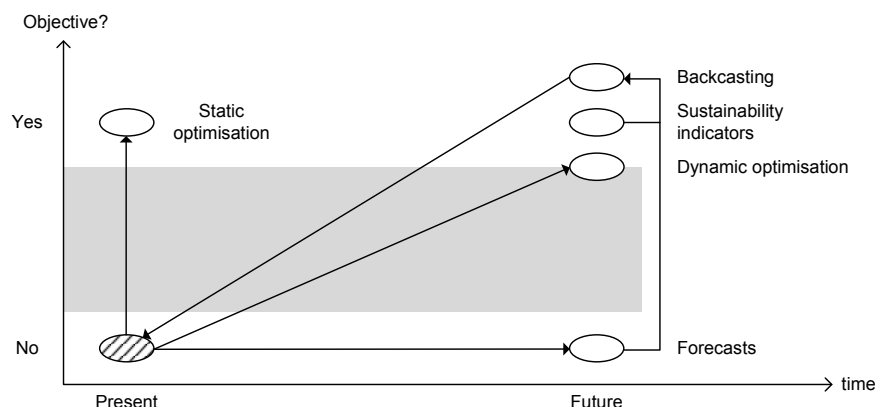


Figure 4.4: Classification of approaches, adapted from Höjer and Mattsson (2000).

Roughly, a distinction can be made between the following four methods and modelling approaches:

1. Sustainability indicators, which are used to indicate (the deviation from) a target value for a certain sustainability variable or combination of sustainability variables;
2. Scenario techniques and backcasting, which are concerned with analysing the future impact of certain scenarios, or backcasting future images (that comply with the objective) to the present situation along transition paths using scenarios;
3. Static optimisation, which is concerned with optimising transport networks for a given network topology and travel demand, using variables as road pricing;
4. Dynamic optimisation, which is used to find transition paths to optimise a transport network over time.

Figure 4.4 illustrates the four different modelling approaches, in relation to traditional forecasting. They are summarised in subsequent paragraphs.

4.3.1 Indicators for transport sustainability

Indicators are central to the monitoring and reporting of progress towards certain public issues. Collectively indicators can measure the capacity to meet present and future needs if it comes to sustainable development or transport sustainability, and therefore has performance-monitoring power. Furthermore, they are often easy to communicate, which is very relevant if it comes to the policy process.

Standard use of indicators to measure and monitor transport performance is done frequently through indicators like level-of-service, traffic speeds, parking

convenience and price, as well as accident rates per vehicle-kilometre. For measuring sustainable transport development, many organisations defined their indicators related to economic, social and environmental sustainability, like total vehicle kilometres, total motorised movement of people and use of fossil fuel energy for all transport, as discussed in length in chapter 2, see for example those of The Centre for Sustainable Transportation (CST, 1997) on page 44, also summarised in Gudmundsson (2001) or Litman (n.d.b).

The problem, however, is how to operationalise the use of single indicators in indicator systems (the collective of (also interacting) indicators), in other words how to measure the ‘overall’ sustainability of a transport system using indicators. In particular, if many different indicator values exist, amongst which some of them even overlap, building indicator systems can become fairly complicated.

To obtain an overall sustainability index multi-criteria analysis (MCA) techniques can be used. Oude Moleman (2001), for example, shows how this can be done in practice for analysing and monitoring sustainable transport development, combining the regime MCA method (using a pairwise comparison of alternatives), as discussed in Hinloopen et al. (1983), within a *Delphi* analysis⁶. The *Compass Index of Sustainability* is then applied, which clusters indicators and assessment scores into different sustainability quadrants. These quadrants are then superaggregated to produce an ‘Overall Sustainability Index’. This index is inspired by the work of Herman Daly (see for example Daly, 1991) and Donella Meadows (see for example Meadows et al., 1972).

Once established, the index provide clear signals about sustainability performance over time compared to an absolute or ‘ideal’ future target.

4.3.2 Scenario techniques and backcasting

In order to explore future developments in transport, to experiment with this future and to evaluate possible effects of policy interventions scenario techniques can be used. Scenario techniques are well-known forecasting techniques, which according to Malone (2001) assist in structuring, understanding and thinking through a changing situation. Furthermore, do they focus attention on the structural uncertainty of the most critical factors, as well as increase the capability of the developers to understand the environment in which the transport system functions. Scenarios contain an integrated description of the future, which pays attention to developments in all factors affecting travel demand, but also provide a plausible sequence of events leading to this situation. In addition, an analysis of the present situation and a connection between future developments and the present situation can be made with such models. An example of the use of scenario techniques can be found in the Dutch SCENARIOEXPLORER model (Malone, 2001).

⁶This is a method for structuring communication in a process that allows a group of individuals to deal with a complex problem and reach consensus.

Backcasting is a variation to the scenario techniques and focuses on presenting solutions to problems that do not seem to be solved, according to conventional scenarios, trends and forecasts. On the basis of a future image that has been identified through a certain vision, for example retrieved by using indicator assessment as discussed in paragraph 4.3.1, or by using scenarios, which may be obtained through a *Delphi* analysis, a path between the future image and today is formulated. Doing so, policy paths towards a certain image are identified, not necessarily being the optimal paths. In comparison to mathematical modelling, backcasting studies come in when forecasting or optimisation studies indicate that sustainability targets cannot be met, as depicted in figure 4.4. Hence, backcasting encourages searching for new paths along which development can take place, particularly when the conventional paths, including those revealed by transport optimisation, do not seem to solve the problem.

Backcasting and forecasting approaches like the scenario approach are complementary. The argument Höjer and Mattsson (2000) put forward is that backcasting is mainly appropriate where current trends are leading towards an unfavourable state, so great non-incremental changes are needed. Forecasting methods remain necessary because they can inform the backcaster when backcasting is required.

Several studies report having used ideas of backcasting for modelling transport sustainability, for example for the Berlin mobility plan in Bluemel (1999), in Höjer (1998) for studying the effect of transport telematics on urban transport and environmentally sustainable transport scenarios, reducing emissions by 80-90%, for the EU in Van Wee and Geurs (2000). These studies have been discussed in Zuidgeest and Van Maarseveen (2000) as well.

An interesting variation to the use of indicators and backcasting has been presented by Miller and Demetsky (1999). They apply regression analysis to determine land-use limitations (the dependent variable) as a function of the capacity of the transport system, loaded to a certain level of service (the independent variable). The capacity of the transport system is measured in terms like number of major through routes in a network, total zonal link volumes or shortest travel distance between zone boundaries and the central business district.

4.3.3 Optimising transport networks

Optimisation of transport systems is often interpreted as the optimisation of transport networks in relation to a fixed or an elastic travel demand. Much literature deals with the so-called network design problem (NDP), see for an extended review Yang and Bell (1998).

The continuous and discrete network design problem can be distinguished. The continuous network design problem (CNDP) takes the network topology as given and is concerned with the parameterisations of the network, as can be

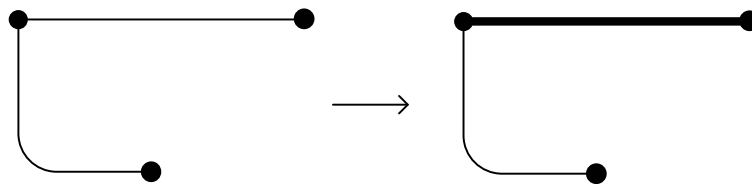


Figure 4.5: Capacity enhancement in the continuous network design problem (CNDP).

seen in figure 4.5. Link parameters that are frequently considered include link (or node) capacity and user charge. CNDP examples are, following Bell and Iida (1999):

1. the determination of road width, that is link capacity or number of lanes;
2. the calculation of traffic signal timings;
3. the setting of user charges (e.g. public transport fares and road pricing).

The DNDP is concerned with the topology of the network, see figure 4.6. DNDP examples are, following Bell and Iida (1999):

1. a road closure scheme;
2. the provision of a new public transport service;
3. the construction of a new road or rail link, a bridge, tunnel or bypass.

Both the CNDP as well as the DNDP are based on the trade-off between user benefit (via travel demand) and the costs of the network alteration (the infrastructure supply). For example the construction of a bus-lane increases the public transport capacity, and therefore public transport throughput, leading to user benefits. However, the construction costs must be set against the user benefits. Furthermore, the user (dis)benefits of the non-public transport users should also be considered. The same applies to the addition or removal of a link in the network compared to the costs of the network alteration. The CNDP and DNDP are sometimes combined in the mixed network design problem (MNDP).

In optimal design of networks there are two sets of interest according to Vickerman (2001). One is the optimal use of the network by the user (i.c. travel demand), who is supposed to be minimising his or her individual generalised costs of travel. The second is the policy maker or transport network operator, who strives to obtain the most efficient transport network (i.c. infrastructure supply) for the given travel demand. Hence, transport networks can be designed that may be optimal to the network operator, but deprives users from their optimal design (i.c. the shortest path in terms of generalised costs). Hence, advanced optimisation techniques, like static bi-level optimisation and dynamic optimisation, have been applied in transport planning to serve both interests in an optimal design.

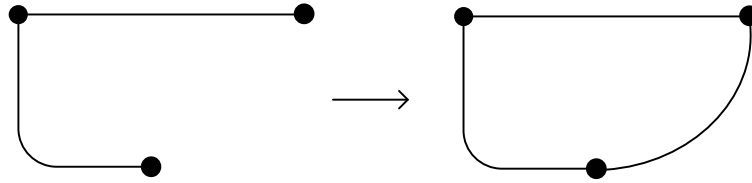


Figure 4.6: Adding a link in the discrete network design problem (DNDP).

Static optimisation

Whereas user optimal and operator optimal network designs have individually been the topic of many researches, for example the travelling salesman problem⁷ in logistics since the 1930s and shortest-path algorithms since the late 1950s, the interaction between the sets of travel demand and infrastructure supply, seen as two interacting levels of optimising agents only got recognition in the early 1970s. Since then the design process is often seen as a bi-level programming problem involving an *upper-level problem* describing the supply problem, and a *lower-level problem* being the demand problem. Here, the designer/operator leads, taking into account how the users follow. The leader has prior knowledge of the responses of the followers, which is also known as a *Stackelberg game* (Stackelberg, 1934), which is in line with both (possibly contradicting) interests of the policy maker versus the user as specified by Vickerman (2001). Bi-level formulations have been applied many times in transport research recently, amongst others in traffic control design (see for example Chiou (1999)), traffic assignment algorithms (see for example Jayakrishnan et al. (1995)), origin-destination table estimation (see for example Yang (1995)), road pricing (see for example Constantin and Florian (1995)) and network design (see for example Yang et al. (2000)).

The general static bi-level optimisation problem can be expressed mathematically, following Yang and Bell (1998), as:

$$\begin{aligned} \min_v \quad & F(u, v(u)) \\ \text{s.t.} \quad & G(u, v(u)) \leq 0, \end{aligned} \tag{4.1}$$

where $v(u)$ is implicitly defined by:

$$\begin{aligned} \min_v \quad & f(u, v) \\ \text{s.t.} \quad & g(u, v) \leq 0, \end{aligned} \tag{4.2}$$

where F and $u \in \mathcal{U}$ are the objective function and decision or control vector for the upper-level decision-maker (the system manager, for example aiming

⁷Given a finite number of nodes in a network along with the cost of travel between each pair of nodes, find the cheapest way of visiting all the nodes in the network and returning to the starting point.

at minimising the total number of kilometres driven in the network), G is the constraint set of the upper-level decision vector, f and v are the objective function and decision vector of the lower-level decision-maker (the traveller, for example aiming at maximising his or her net-utility through route choice decisions), and g is the constraint set of the lower-level decision vector. Hence, the lower-level of the bi-level model represents the demand-performance equilibrium for a given investment action by the user. The whole bi-level NDP is to find the optimal investment decision-making, for example through optimal capacity improvement u^* such that the system objective function F is optimised subject to a given budget constraint in the upper-level program of equation (4.1), while taking account of the reaction of the network users in the lower-level program of equation (4.2).

Applying the static bi-level optimisation techniques to derive sustainable transport network can be seen as having an upper-level sustainability goal, while complying with the lower-level behavioural rules of the travellers in the system. An example of such applications can be found in Yang and Bell (1997), where road pricing is being advocated as an efficient means of managing traffic demand, while complying with capacity constraints representing other objectives, such as reducing the environmental impact of road traffic and improving public transport. In addition, Kim (1998) presents for example bi-level formulations for modelling integrated urban land-use - transport interaction.

Apart from bi-level techniques, network equilibrium techniques, using *variational inequality* theory, are used to model sustainable transport networks. Variational inequality theory provides a framework for the study of one-level equilibrium problems, particularly for modelling Wardrop's First Principle (see paragraph 1.2.2). For modelling the sustainable transport optimisation problem, variational inequality theory has been extensively applied by Anna Nagurney, who considers a transport network to be sustainable if 'the flow pattern satisfies the conservation of flow equations and does not exceed the imposed environmental quality standard, subject to the operating behavioural principle's underlying the network' (Nagurney, 2000). Transport policies are meant to direct transport networks to sustainability (assuming viability, which implies environmental goals are achievable, given the network, travel demand and environmental parameters). Her policies include emission pricing, introduction of tradable pollution permits as well as introduction of new modes. Variational inequality theory is also used by Paolo Ferrari who also models sustainable transport networks. In Ferrari (1995) a deterministic variational inequality equilibrium model for urban transport networks with elastic demand and two types of capacity constraints is presented. These capacity constraints can be physical and environmental. More recently, in Ferrari (1999), a similar study is presented, modelling a multimodal transport system, subject to some physical and environmental capacity constraints, as well as budget constraints.

The concept of a sustainable and developing transport systems as defined in paragraph 2.3.1, applying the concept of productive capacity from equa-

tion (3.7), has been implemented in the Urban Traffic Environmental analysis Model (UTEM). The model has been used extensively for optimising artery's in several cities, as reported in Zuidgeest et al. (2001) as well as Akinyemi and Zuidgeest (2002). Supply-side interventions are generated iteratively in UTEM to maximise an accessibility criterion based on the productive capacity of the network and generalised costs of travel, while complying with social, environmental and financial resource capacities.

In addition, applying mathematical programming techniques, Huapu and Peng (2001) discuss a static constrained optimisation model for sustainable urban transport planning. Their model maximises trip making, constrained by an environmental capacity parameter for exhausted gasses as well as an ecological footprint capacity⁸.

Dynamic optimisation

The 'traditional' equilibrium models discussed before assume traffic in the optimal network design is in a static user-equilibrium, and that changes through decision variables (for example physical capacity, road pricing etceteras) do not induce transient effects. However, in view of the characterisation of sustainable urban transport development, a time-varying nature of travel demand and infrastructure supply in the network seems obvious. Hence, dynamic optimisation models have been introduced that take into account both the time-varying nature of transport networks as well as disequilibrating effects that changes to the network may produce, hence reflecting time-lags in adaption processes towards the equilibrium put forward in the static models. Furthermore, dynamic modelling of transport is said to have some additional advantages over static modelling as the absence of the so-called temporal *Braess paradox*, which says that sometimes in transport networks paradoxical behaviour may appear, that is when an enhancement of road capacity results in worse traffic conditions for all road users (Bell and Iida, 1999). Furthermore, the unrealistic sequential nature of the traditional transport planning model and the lack of feedback seem better treated by dynamic versions of this model (as recalculations of the traditional models occur each time-step). In addition, dynamic optimisation allows for determining the paths of control variables and state variables for the dynamic disequilibrium system over a finite or infinite time horizon to maximise an objective function. Hence, dynamic optimisation is also called optimal control.

Dynamic optimisation models for optimising transport networks have first been published by Friesz and Fernandez (1979), who developed a linear optimal control problem in order to determine optimal maintenance policies, knowing that the quality of a road facility is determined by natural factors, rate of use and maintenance investments, whereas the demand for the road facility is again

⁸Ecological footprint is an index of sustainable development based on land-use area. It quantifies the intensity of source consuming related with the regional bearing capability. A famous introduction in the concept is given in Wackernagel and Rees (1995).

assumed to be a function of its quality under these circumstances. An early example of optimising transport networks is given in Steenbrink (1974), who applied heuristic iterative optimisation methods to derive bi-level optimisation problems in transport network design. Nijkamp and Reggiani (1988) derived a dynamic model that represents the evolution of an integrated spatial interaction system. Friesz and Shah (1998) discuss the theory of disequilibrium network design and use this in a model to maximise social welfare in a transport system. Donaghy and Schintler (1994) *ibid.* Donaghy and Schintler (1998), as does Lensink (2002), study the possibility of using optimal control techniques for modelling infrastructure management in terms of road capacity enhancements, and road/parking pricing. On a strategic level the European Commission 'Optimisation of policies for transport integration in metropolitan areas (OPTIMA)' study, as reported by European Commission (1998) *ibid.* Pfaffenbichler and Shepherd (2002) or Pfaffenbichler (2003), tested several (pre-defined) packages of sustainable transport measures in a dynamic transport model against two objective functions (an economic efficiency function as well as a sustainability objective function), hence finding optimal policy packages applying multidimensional minimisation techniques, instead of optimal control techniques.

In the sustainable urban transport development concept presented in this research a methodology is proposed that uses a standard forecasting framework in which goals are explicitly defined. Hence, it operates partly in the highlighted area of figure 4.4. Backcasting techniques should be applied when standard transport policy measures cannot somehow lead to the desired targets. In other words, if non-incremental changes, perhaps introducing new means of transport, are needed.

Next chapter discusses dynamic optimisation in more detail and will introduce an optimal control model based on the concept of sustainable urban transport development. This model can be used as a tool for designing sustainable transport policies.

Chapter 5

A dynamic optimisation model for sustainable urban transport development

5.1 Introduction

Optimisation is part of everyday life. For example, when dealing with scarce resources, as road space or environmental capacity, one would like to optimise the use of it. Several optimisation techniques exist for different types of optimisation problems (Naidu, 2003); for example static optimisation problems, when confronted with controlling a system under steady-state conditions, game theoretic problems, when multiple actors are optimising individually within one system, dynamic programming problems, when dealing with multiple stage problems (preferably in discrete-time) in the sense that at each set of times, a decision can be chosen from a finite number of decisions based on some optimisation criterion, or stochastic problems, where the nature of the signal is stochastic, etceteras. Deterministic, single stage optimisation problems in continuous-time can be treated as dynamic optimisation problems. Dynamic optimisation is concerned with the optimal control of systems under dynamic conditions, i.e. the system variables are changing with respect to time and thus time is included in the differential or difference equations describing the system.

In this chapter, dynamic optimisation is used to model sustainable urban transport development in accordance with the characterisations and theory developed in previous chapters. First, some basics on dynamic optimisation are given. Followed by a detailed description of the dynamic transport model that is solved applying the Maximum Principle of Pontryagin, which is discussed and demonstrated accordingly.

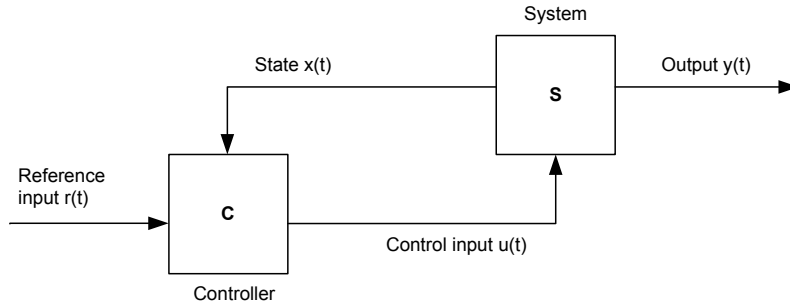


Figure 5.1: The general control configuration.

5.2 Dynamic optimisation

In general control theory a controller C is determined that will cause the (non-linear) dynamic system $\mathcal{S}(\mathcal{T}, \mathcal{W}, \mathcal{B})$, as characterised in paragraph 4.2, which is described in terms of state equations $x(t)$ to satisfy some physical constraints and at the same time maximise or minimise a chosen performance index J . The system output $y(t)$ is controlled through the control input $u(t)$, which is driven by the state equations and some reference signal or input $r(t)$, which is illustrated in figure 5.1 as a so-called closed-loop control.

In *optimal control*, a control $u^*(t)$ is looked for that will drive the system \mathcal{S} from the initial state (at time t_0) to a final state (at time t_1) with some constraints on the controls and states, while at the same time minimising or maximising the given performance index J , which is some function of the system output $y(t)$ as well as controls $u^*(t)$ over time.

The formulation of the optimal control problem therefore requires (Naidu, 2003):

1. a mathematical description of the system to be controlled;
2. a specification of the performance index J ;
3. a statement of boundary conditions and physical constraints on the states and/or controls.

If a nonlinear ordinary differential system with initial conditions for the state variable $x(t)$, also equation of motion or transition, is considered:

$$\frac{dx(t)}{dt} = f(x(t), u(t), t), \quad x(t_0) = x_0, \quad (5.1)$$

with the control variable $u(t)$, which like the state variable $x(t)$, may be bounded by some equality or inequality constraint:

$$g(x(t), u(t), t) \geq 0, \quad (5.2)$$

then the performance index, or cost criterion or functional, J , can be formulated as a so-called *Bolza-problem*¹:

$$J = F(x(t_1), t_1) + \int_{t_0}^{t_1} H(x(t), u(t), t) dt, \quad (5.3)$$

where $x(t)$ is the solution of equation (5.1) at time t , F is some final state, which may be fully, partially fixed or free; also called transversality condition. Furthermore, there is some (piecewise) continuous function H . The final time t_1 may be fixed or free also.

The optimal control problem for a minimisation problem can be stated as the determination of the control function $u^*(t)$, given system equation (5.1), constraints (5.2) and cost criterion (5.3), such that:

$$J^* = J(x_0, u^*(t), t) \leq J(x_0, u(t), t), \quad \forall u(t). \quad (5.4)$$

Dynamic programming (sometimes referred to as Bellman's Optimality Principle) and Pontryagin's Maximum Principle are two basic tools for studying optimal control theory, and can be found in any textbook on optimal control. Dynamic programming determines a closed-loop optimal control that is optimal everywhere, whereas the open-loop Maximum Principle refers only to quantities along a specific trajectory. However, to obtain the optimal control using dynamic programming one has to discretise a high dimensional state space, which causes an exponential growth of the number of grid points with dimensionality (referred to as the curse of dimensionality). Computational methods based on Pontryagin's Maximum Principle avoid this curse of dimensionality. Given the very large dimensionality of transport network state spaces, it is believed here that computational methods based on the Maximum Principle are more promising than the methods based on dynamic programming, and is henceforth applied here. First, the nature of the applied model equations as well as the dynamic transport model itself are discussed, before turning to a description of the Maximum Principle in paragraph 5.4.

5.2.1 Continuous-time modelling

Dynamic optimisation techniques can be formulated in continuous-time as well as discrete-time. In discrete-time, a period of time $t \in \mathbb{N}$ is divided into n intervals of length T , so time is indexed as x_t . In continuous-time $t \in \mathbb{R}$ in the interval $[t_0, t_1]$, so a variable is a function of time $x(t)$. Although the world is

¹A number of variations to general class of Bolza-problems exist, for example the Mayer-problem if $H = 0$, or Lagrange-problem if $F = 0$ in equation (5.3).

observed in discrete-time only, it is often much easier to work with continuous-time.

A few main arguments are postulated in favour of the use of continuous models:

1. Although individual strategic planning decisions are generally made at discrete time-intervals, it is difficult to believe that they are co-ordinated in such a way as to be perfectly synchronised; they will overlap in time in some way. As the variables that are usually considered and observed are the outcome of a great number of decisions taken by different operators at different points in time, it seems natural to treat the different processes as if they were continuous (Gandolfo, 1993). This also takes away the need to make a decision on what is the appropriate ‘natural’ unit of time, where time is expressed in for example: $T = 1$ [month], or: $T = 1$ [quarter];
2. A specification in continuous-time is particularly useful for the modelling of disequilibrium adjustment processes like for example the inducement of travel demand (see chapter 3);
3. The availability of a model formulated as a system of differential equations enables its user to get forecasts and simulations for any time-interval, and not only for the time unit inherent in the data;
4. Continuous-time specifications are more tractable mathematically and more consistent with growth-theory frameworks and diffusion models. Hence, they also provide a better way of depicting ongoing aggregate behaviour (Donaghy and Schintler, 1998);
5. The state of transport systems is one which is continuously evolving (see chapter 1);
6. Development in itself is a dynamic, continuous phenomenon, therefore also the strategic development of a transport system (see chapter 2).

In this research a continuously evolving state of the transport system is modelled that can be observed at regular intervals. Therefore the state of the transport system can be determined at these time intervals.

5.2.2 Disequilibrium modelling

Optimising agents after calculating the optimal value of some decision variable, cannot immediately adjust the actual to the desired value due to frictions and imperfections of various types (Gandolfo, 1997). Hence, it is questionable whether a static equilibrium, i.e. ‘a transport system finds itself in an essentially timeless state where system users are either unable or unwilling to change their behaviour’ (Bell and Allsop, 1998), exists. At best a transport system is near equilibrium, or tending towards it, only prevented by external factors. To include time-lagged processes, as reaction times or habitual chances, partial adjustment equations can be used. In a partial adjustment equation, indicating

a discrepancy between the desired and actual value of a variable, this process can be modelled. The actual value is then adjusted towards the desired value only gradually, according to a coefficient of reaction or speed of adjustment.

Disequilibrium modelling can be illustrated using an equation in discrete-time, following Gandolfo (1997). When the value of y at time $t \in \mathbb{N}$ depends on the present and past values of some other variable x , a distributed-lag equation is used, which in discrete time (n time periods) can be written as:

$$y_t = b_0x_t + b_1x_{t-1} + \cdots + b_nx_{t-n}, \quad (5.5)$$

where b_0, b_1, \dots, b_n are known nonnegative constants, with:

$$\sum_{i=0}^n b_i = b. \quad (5.6)$$

Often this series b_i is declined geometrically for some: $0 < k < 1$, as in:

$$b_i = kb_{i-1}. \quad (5.7)$$

Hence, equation (5.6) changes to:

$$b = \sum_{i=0}^{\infty} b_i = b_0 \frac{1}{1-k}. \quad (5.8)$$

The equivalent continuous-time formulation to equation (5.5) is:

$$y(t) = \int_0^{\infty} [f(t')x(t-t')] dt', \quad (5.9)$$

which is a convolution equation, also called a continuously distributed-lag equation. The function $f(t')$ is called a weighting function, which may be:

$$\int_0^{\infty} f(t') dt' = \frac{1}{b}, \quad (5.10)$$

or an exponential weighting function, analogue to equation (5.7):

$$f(t) = \gamma e^{-\gamma t}, \quad (5.11)$$

hence, equation (5.9) can be written as:

$$y(t) = \int_0^{\infty} \gamma e^{-\gamma t'} x(t-t') dt'. \quad (5.12)$$

It can be shown that this equation is equivalent to a partial adjustment differential equation in continuous-time, i.e.:

$$\frac{dy(t)}{dt} = \gamma(x(t) - y(t)). \quad (5.13)$$

For a proof, the reader is further referred to Gandolfo (1997).

The coefficient of adjustment γ can be interpreted as the reciprocal of the mean time-lag, which is the time required for about 63% of the discrepancy between $y(t)$ and $x(t)$ to be eliminated by changes in $y(t)$, following a change in $x(t)$. The mean time-lag is defined as:

$$\bar{t} = \int_0^{\infty} t' f(t') dt', \quad (5.14)$$

which is equivalent to:

$$\bar{t} = \gamma \int_0^{\infty} t' e^{-\gamma t'} dt' = \frac{1}{\gamma}, \quad (5.15)$$

for the exponential weighting function: $f(t) = \gamma e^{-\gamma t}$. So, when: $\lim \gamma \rightarrow \infty$, the mean time-lag tends to zero, which means that $y(t)$ adjusts to $x(t)$ immediately.

The partial adjustment equation (5.13) can be used in the specification of dynamic disequilibrium models. In this type of models, it is claimed that the dynamic system is continuously adjusting to the partial equilibrium $x(t)$, for example due to some positive external influence F^+ , but effectively is in a disequilibrium $y(t)$, except when the system is in a steady-state, which is shown in figure 5.3. This is different from an instantaneous adjustment as depicted in figure 5.2. In terms of transport, this can be seen as the time lagged adjustment of travel demand to changes in exogenous social-economic factors, but also to changes in infrastructure supply. Referring to Manheim's equilibrium, as depicted in figure 1.5, the equilibrium point is aimed for, but due to the system dynamics and time-lags the transport system is said to be in a *disequilibrium state* continuously. As time is passing several positive and negative external influences F will change the disequilibrium, which is illustrated in figure 5.4. Basically, the disequilibrium process guides the transport system from one partial equilibrium state to another, perhaps eventually reaching a steady state, the static equilibrium.

Disequilibrium transport network design models have not been used widely. As said before, applications can be found in transport network design, i.e. Friesz and Shah (1998), Donaghy and Schintler (1994) *ibid.* Donaghy and Schintler (1998).

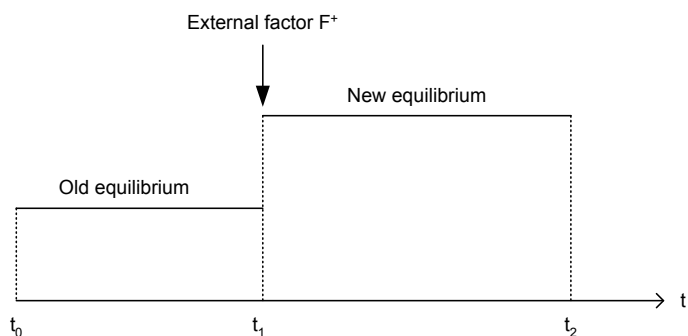


Figure 5.2: Instantaneous adjustment to an external factor F^+ (De la Barra, 1989).

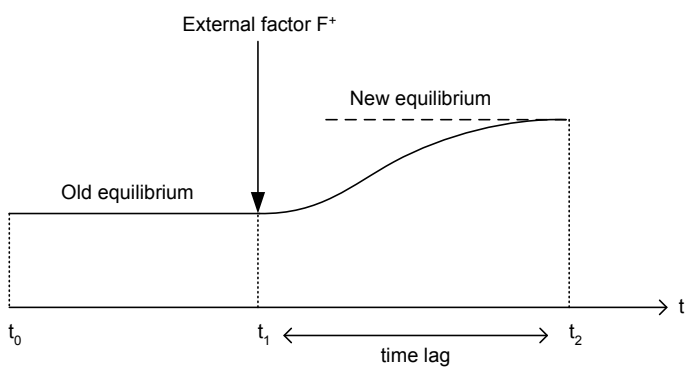


Figure 5.3: Dynamic lagged adjustment to an external factor F^+ (De la Barra, 1989).

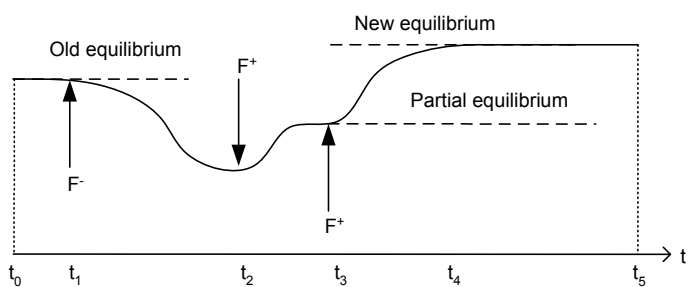


Figure 5.4: Dynamic lagged adjustment to several external factors F .

5.3 A dynamic transport optimisation model

In this paragraph an optimal control model for sustainable urban transport development is derived. The equations are formulated as such, to prepare for the use of the Pontryagin's Maximum Principle, which is the topic of paragraph 5.4. The dynamic model is based on the conceptual ideas set forward in chapter 2, where sustainable urban transport development has been defined, and chapter 4, where sustainable transport planning has been characterised as an optimisation problem. The basic model equations are based on the traditional transport planning model equations from chapter 1 as well as the sustainability requirements from chapter 3. Summarising, the dynamic model can be characterised by its:

Dynamics The transition paths in time towards a sustainably developed transport system are described by state variables representing traffic performance (flow, capacity, condition) and impacts (emissions, throughput, kilometres travelled) per road link in the network as well as the costate variables with these state variables representing the marginal costs;

Optimisation The nature of the problem of designing a sustainable transport planning is that of a constrained optimisation problem. Different transport policy objectives are translated into cost functions over time (the planning horizon), i.e. maximising accessibility, maximising person throughput or minimising vehicle congestion, which can be studied and compared;

Controls Transport improvements are related to travel time savings and vehicle operating costs savings, hence affecting generalised costs of travelling. The improvements can be effected through controls to the decision maker: engineering interventions (road construction, road maintenance, public transport priority) as well as pricing measures (vehicle taxes, parking taxes, bus fares);

Constraints The optimisation and use of controls are limited by (endpoint) constraints, i.e. environmental and financial capacities the decision maker has to comply with at the end of the planning horizon, but also constraints that physically limit the control value or state value over time.

The basic idea behind this dynamic optimisation model is that a transport planner who, confronted with some transport policy objective as well as several (endpoint) constraints or resource capacities (like infrastructure budget, environmental capacity and equity objectives), has at stake, different (combinations of) transport measures or controls, like new construction, road maintenance as well as pricing that will change the effective road capacity and productive capacity of the network, hence affecting traffic performance. The composite generalised costs, consisting of vehicle-operating-costs as well as travel time for the different modes on different routes, will influence accessibility, which is also dependent on the interaction potential between zones (the trip generation capacity). From this the total elastic mode-specific travel demand, and

consequently traffic flows, can be determined on the basis of the equilibration process of travel demand and infrastructure supply. Available resources are consumed from the resource capacities available. This general structure of the dynamic transport model can be depicted as a diagram with causal relations, see figure 5.5.

Among the transport system attributes in the model are things as route specific travel times, travel costs and capacities (at some reference condition or during the period when traffic is heaviest, i.e. during peak hours). These are attributes observed at points on a continuum of time, necessary for medium and long term strategic transport planning purposes, not to be confused with short-term dynamics of (for example) non-recurrent traffic congestion.

The basic idea behind the optimal control models discussed here are inspired by the work of Laurie Schintler and Kieran Donaghy, as published in Donaghy and Schintler (1994) *ibid.* Donaghy and Schintler (1998). Similar models have been reported in Büttler and Shortreed (1978), Friesz and Fernandez (1979) and Lensink (2002).

In the next paragraphs, the specific elements of the dynamic model are discussed, subdivided into dynamic models for travel demand, infrastructure supply and system induced effects in paragraph 5.3.1, transport planners objective functions in paragraph 5.3.2 as well as constraints in paragraph 5.3.3.

5.3.1 Transport dynamics

In the optimal control model, transport dynamics are expressed using a set of ordinary differential equations. The main state variables $x(t)$ are travel demand $V(t)$, using a disequilibrium formulation as discussed in paragraph 5.2.2, as well as infrastructure supply, or link capacity $C(t)$. Furthermore, some other state variables representing the transport induced effects, i.e. vehicle-kilometres travelled $K(t)$, person- throughput $P(t)$ and total emissions $E(t)$ are formulated. Each state variable is discussed subsequently.

The spatial structure of the transport network is represented using subscript indices with the different variables. Let the graph $\mathcal{G}(\mathcal{N}, \mathcal{L})$ be a transport network with nodes $n \in \mathcal{N}$ and links $l \in \mathcal{L}$. The link set is: $\mathcal{L} = \{1, 2, \dots, l, \dots, L\} \in \mathbb{N}^{\mathcal{L}}$. Each link may have several attributes, such as length, free-flow speed, capacity, road type etceteras. The centroid-set: $\mathcal{Z} = (\mathcal{I}, \mathcal{J}) \subset \mathcal{N}$, is a subset of \mathcal{N} and consists of origins i , numbered: $\mathcal{I} = \{1, 2, \dots, i, \dots, I\} \in \mathbb{N}^{\mathcal{I}}$, as well as destinations j , numbered: $\mathcal{J} = \{1, 2, \dots, j, \dots, J\} \in \mathbb{N}^{\mathcal{J}}$. Similarly, a mode-set can be defined as: $\mathcal{M} = \{1, 2, \dots, m, \dots, M\} \in \mathbb{N}^{\mathcal{M}}$, whereas the route-set between origin-destination pair (i, j) is: $\mathcal{R}_{ij} = \{1, 2, \dots, r, \dots, R\} \in \mathbb{N}^{\mathcal{R}}$. Social-economic population segmentation is equally represented in a set: $\mathcal{K} = \{1, 2, \dots, k, \dots, K\} \in \mathbb{N}^{\mathcal{K}}$. A pollutant set: $\mathcal{P} = \{1, 2, \dots, p, \dots, P\} \in \mathbb{N}^{\mathcal{P}}$, is defined as well. Finally, a control set: $\mathcal{U} = \{1, 2, \dots, u, \dots, U\} \in \mathbb{N}^{\mathcal{U}}$, contains all possible control options.

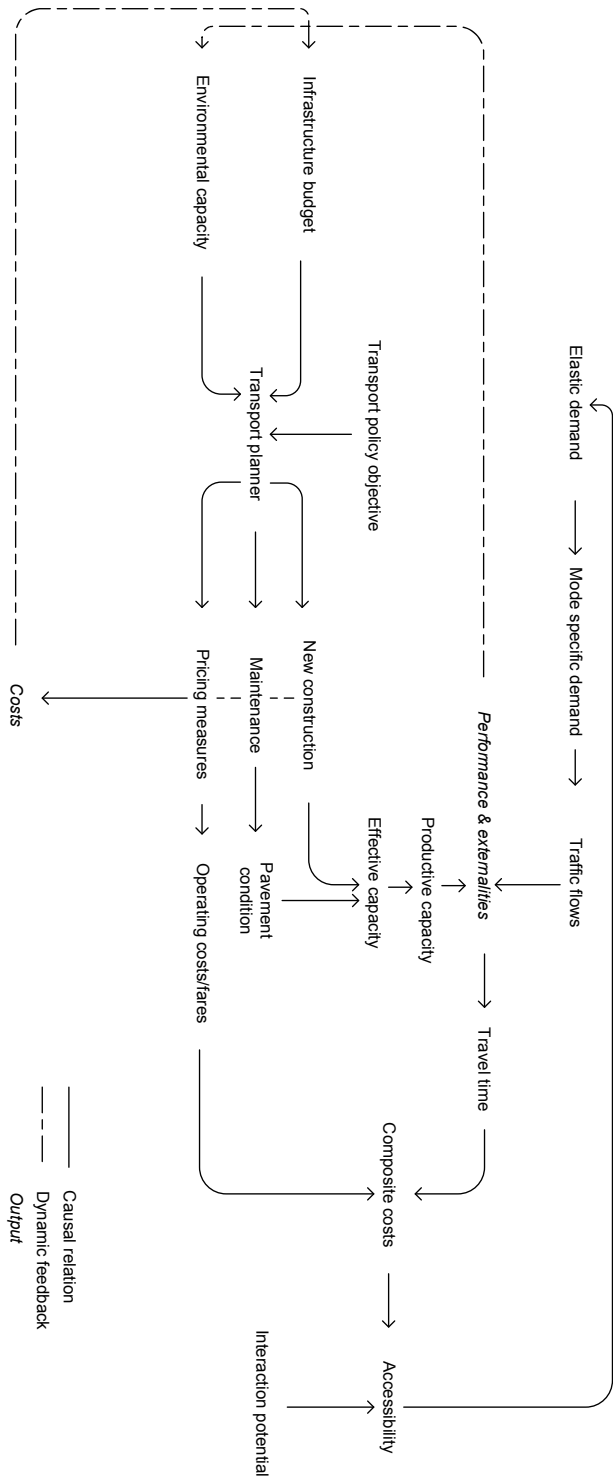


Figure 5.5: Causal structure of the dynamic transport optimisation model, adapted from Donaghy and Schintler (1998).

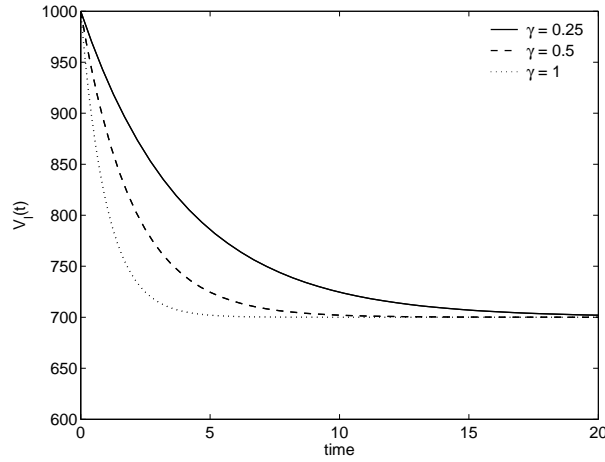


Figure 5.6: The changing state of travel demand for different values of γ_1 .

A disequilibrium travel demand model with static trip generation capacity (A1)

Disequilibrium equation (5.13) can be used to model the changing state of travel demand $V(t)$ on a road link over time t . If the trip generation capacity Q is fixed (for simplicity Q is denoted here in $[\text{pcuh}^{-1}]$ instead of $[\text{pers h}^{-1}]$), the dynamic travel demand model can be written as a first-order ordinary differential equation:

$$\frac{dV(t)}{dt} = \gamma_1 (Q - V(t)). \quad (5.16)$$

The solution to this equation obviously is:

$$V(t) = Q + (V_0 - Q)e^{-\gamma_1(t-t_0)}, \quad (5.17)$$

where V_0 is the initial state for $V(t)$ at: $t = t_0 (= 0)$. The solution² is graphically depicted in figure 5.6 for different values of the adjustment coefficient γ_1 , while: $Q = 700 \text{pcuh}^{-1}$ and: $V_0 = 1000 \text{pcuh}^{-1}$. For relatively small relaxation time, implying high γ_1 the disequilibrium system moves quickly from V_0 to Q .

A disequilibrium travel demand model with time-varying trip generation capacity (A2)

Exogenous explaining (or independent) social-economic or land-use variables might influence the dependent trip generation capacity Q over time, hence

²A variable-step continuous solver ODE45 based on an explicit Runge-Kutta (4,5) formula is used in MATLAB.

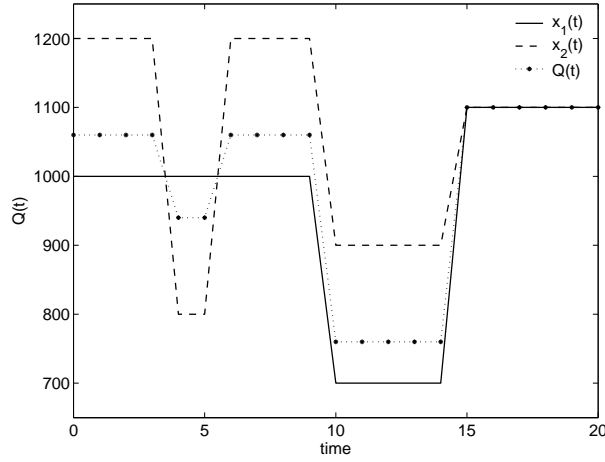


Figure 5.7: Dependent $Q(t)$ and independent exogenous variables $x_1(t)$, $x_2(t)$, parameters: $\alpha_1 = 0.7$, $\alpha_2 = 0.3$, and: $\alpha_3 = 0$.

creating a time-varying $Q(t)$, which can be calculated by constructing a linear multiple-regression equation:

$$Q(t) = \sum_{k_1=1}^{n_{k_1}} (\alpha_{k_1} x_{k_1}(t)) + \alpha_{k_1+1}, \quad (5.18)$$

with in total n_{k_1} exogenous explaining variables in vector $\vec{x}(t)$, with parameter vector $\vec{\alpha}$ and α_{k_1+1} as the intercept. An example of $Q(t)$ is given in figure 5.7, with time-varying (arbitrary chosen) explaining variables $x_1(t)$ and $x_2(t)$, for example representing the number of households as well as average monthly income. The values of x_1 and x_2 vary over time, and therefore the dependent variable $Q(t)$ varies accordingly.

Now the dynamic travel demand equation (5.16) changes to:

$$\frac{dV(t)}{dt} = \gamma_1 (Q(t) - V(t)), \quad (5.19)$$

of which the solution is:

$$V(t) = e^{-\gamma_1(t-t_0)} \left(V_0 + \gamma_1 \int_{t_0}^{t_1} Q(t') e^{\gamma_1 t'} dt' \right), \quad (5.20)$$

with V_0 the initial state for $V(t)$ at $t = t_0 (= 0)$. The solution is graphically depicted in figure 5.8 for different values of the adjustment coefficient γ_1 , while: $V_0 = 1000 \text{pcuh}^{-1}$. It can be observed that for a high: $\gamma_1 = 10$, implying small relaxation time, the exogenous $Q(t)$ from figure 5.7 is retrieved.

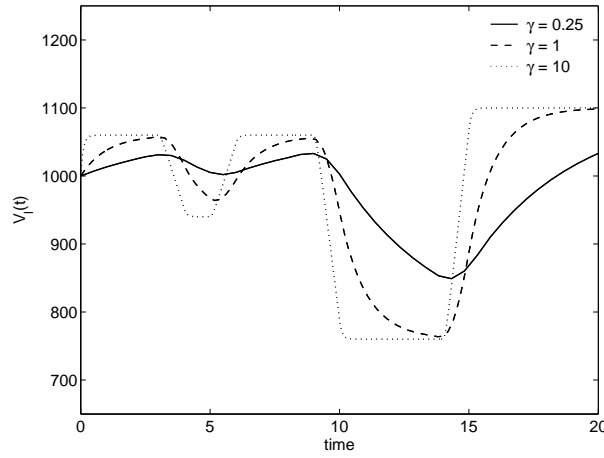


Figure 5.8: The changing state of travel demand for different values of γ_1 .

A disequilibrium network travel demand model with time-varying trip generation capacity (A3)

In a transport network, link travel demand is dependent on the choice-behaviour of the trip maker. Furthermore, several origins, destinations, modes and routes exist. Therefore, the time-varying trip generation capacity in an origin-zone i is distributed over several destinations j , modes m and routes r . To capture this choice-behaviour the four-step model equations from paragraph 1.2 can be used to replace $Q(t)$ in equation (5.19) and find the equilibrium traffic flow $\hat{V}_l(t)$ for link l at time t .

Doing so, a disequilibrium formulation, following Donaghy and Schintler (1998), expresses the changing state of travel demand (in $[\text{pcuh}^{-1}]$) between the equilibrium traffic volume $\hat{V}_l(t)$ and actual traffic volume $V_l(t)$. Depicted again as a dynamic travel demand model this reads:

$$\frac{dV_l(t)}{dt} = \gamma_1 \left(\hat{V}_l(t) - V_l(t) \right), \quad \forall l. \quad (5.21)$$

The equilibrium link demand function $\hat{V}_l(t)$ is assumed to be built-up of:

1. the time-varying trip generation capacity of a zone i , based on social-economic characteristics of the zone, $Q_i(t)$;
2. a simultaneous mode m , destination j and route r choice model, $G_{ijmr}(t)$.

The generated trips $Q_i(t)$ in zone i are distributed over the different choice-options using a discrete-choice model $G_{ijmr}(t)$ and converted to vehicles in $[\text{pcuh}^{-1}]$ applying a vehicle-occupancy factor θ_{1m} :

$$\hat{V}_l(t) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^l} \theta_{1m} (Q_i(t) G_{ijmr}(t)), \quad \forall l, \quad (5.22)$$

with \mathcal{R}_{ij}^l the route serving origin-destination pairs that contain link l , or: $\mathcal{R}_{ij}^l = \{r \in \mathcal{R}_{ij} \mid \exists n : r_n = l\}$, and: $r = (1, 2, \dots, l_r, \dots, L_r) \in \mathbb{N}^{N_r}$, as well as \mathcal{J}^i the destination set excluding destination $j = i$, or: $\mathcal{J}^i = \{j \in \mathcal{J} \mid \nexists n : j_n = i\}$.

The trip generation per zone is expressed as a linear multiple-regression equation, being a slight variation to equation (5.18):

$$Q_i(t) = \sum_{k_1=1}^{n_{k_1}} (\alpha_{k_1 i} x_{k_1 i}(t)) + \alpha_{k_1+1 i}, \quad \forall i. \quad (5.23)$$

The chances of simultaneously selecting a certain destination j , mode m and route r are obtained using a dynamised version of the doubly-constrained gravity model of equation (1.19), which is distributing trips on the basis of the utilities $u_{ijmr}(t)$ for the different mode m , and route r choice-combinations as well as the time-varying trip attraction value $X_j(t)$:

$$G_{ijmr}(t) = \frac{X_j(t) \exp(-\lambda_1 u_{ijmr}(t))}{\sum_{j' \in \mathcal{J}^i} \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij'}} X_{j'}(t) \exp(-\lambda_1 u_{ij'm'r'}(t))}, \quad \forall i, j, m, r, \quad (5.24)$$

with scale parameter λ_1 .

The destination attractiveness $X_j(t)$ is equally expressed as the trip generation capacity:

$$X_j(t) = \sum_{k_2=1}^{n_{k_2}} (\alpha_{k_2 j} x_{k_2 j}(t)) + \alpha_{k_2+1 j}, \quad \forall j, \quad (5.25)$$

with n_{k_2} the number of exogenous explaining variables in vector $\vec{x}_k(t)$, with parameter vector $\vec{\alpha}_k$ and $\alpha_{k_2+1 j}$ as the intercept.

The utility function $u_{ijmr}(t)$ for generalised costs of travel is expressed as a linear equation of route travel time $\tau_r(t)$ and route travel cost $\kappa_{ijmr}(t)$ ³, similar to equation (1.1):

³Often, but not here for reasons of simplification, travel costs are weighted to an income parameter if a population segmentation $k \in \mathcal{K}$ is considered, hence the second term in equation (5.26) changes to $\beta_{4m} \frac{\kappa_{ijmr}(t)}{\alpha_2 \pi_k}$, with π_k the income-level for segmentation k and α_2 some kind of *time-value-of-money* parameter, see for example Odoki et al. (2001).

$$u_{ijmr}(t) = \beta_{3m}\tau_r(t) + \beta_{4m}\kappa_{ijmr}(t), \quad \forall i, j, m, r, \quad (5.26)$$

with β_{3m} and β_{4m} mode-specific parameters. Parameter β_{3m} is the *value-of-time* parameter for conversion of time to monetary units. Similarly β_{4m} can be expressed as the *time-value-of-money* parameter for conversion of money to time units.

Route travel time is a summation of link travel times that comprise route r , which are expressed as strictly increasing, continuous and nonlinear functions of the volume $V_l(t)$ to effective capacity $C_l^e(t)$ ratio and free-flow travel time τ_l^0 , by applying equation (1.17) :

$$\tau_r(t) = \sum_{l_r=1}^{N_r} \tau_{l_r}^0 \left[1.0 + \alpha_1 \left(\frac{V_{l_r}(t)}{C_{l_r}^e(t)} \right)^{\beta_1} \right], \quad \forall r, \quad (5.27)$$

with parameters α_1 and β_1 .

Route travel costs are similarly expressed as a function of the exogenous time-varying vehicle operating costs o_m as well as distance dependent vehicle tax control $U^v(t)$ and a parking tax control at the destination $U_j^p(t)$ (a flat tariff is assumed)⁴:

$$\kappa_{ijmr}(t) = (o_m + \theta_{2m}U^v(t))d_r + \theta_{2m}U_j^p(t), \quad \forall i, j, m, r, \quad (5.28)$$

with parameter θ_{2m} representing the conversion of imposed taxes to the vehicle-type m , in other words the disutility per vehicle-type.

The route distance is accordingly calculated as:

$$d_r = \sum_{l_r=1}^{N_r} d_{l_r}, \quad \forall r, \quad (5.29)$$

where d_l is the length of link l , and N_r the number of links comprising route $r \in \mathcal{R}_{ij}$.

The solution to this model now is:

$$V_l(t) = e^{-\gamma_1(t-t_0)} \left(V_{l0} + \gamma_1 \int_{t_0}^{t_1} \hat{V}_l(t') e^{\gamma_1 t'} dt' \right), \quad \forall l, \quad (5.30)$$

⁴In fact, the subscript i could be omitted here as it is not part of the function itself, because each r is determined by a unique combination of i , j , and m . For reasons of clarity the utility and alike remain, however, indicated with subscripts i, j, m, r .

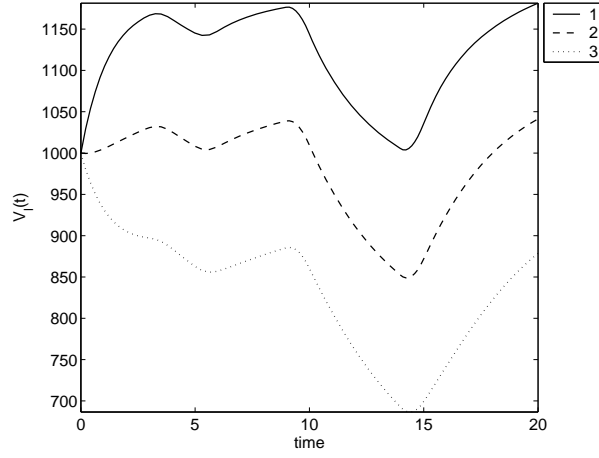


Figure 5.9: The changing state of travel demand in a small network.

with V_{i0} the initial state for $V_i(t)$ at: $t = t_0 (= 0)$. This solution is illustrated in figure 5.9 for a three-link network (assuming one origin, one destination, and one mode), where the, fixed, effective link capacity: $C_1^e(t) < C_2^e(t) < C_3^e(t)$, hence activating the route-choice r only. The adjustment coefficient is: $\gamma_1 = 0.25$, while the initial states are: $V_1(0) = V_2(0) = V_3(0) = 1000 \text{pcuh}^{-1}$. The time-varying exogenous trip generation capacity $Q_i(t)$ is the same as in previous example (see figure 5.7).

A disequilibrium network elastic demand model with time-varying trip generation capacity (A4)

If travel demand is regarded to be elastic to changes in accessibility, as discussed before in paragraph 3.3.1, the equilibrium level $\hat{V}_i(t)$ should also include a measure of elasticity.

This implies that not necessarily all trip generation capacity (still based on the social-economic and land-use characteristics of the zone) is revealed per se. In other words, the *potential* trip generation capacity of a zone needs to be known.

Starting with the disequilibrium travel demand model:

$$\frac{dV_i(t)}{dt} = \gamma_1 \left(\hat{V}_i(t) - V_i(t) \right), \quad \forall i, \quad (5.31)$$

the equilibrium link demand function $\hat{V}_i(t)$ is now assumed to be built-up of:

1. the time-varying potential trip generation capacity of a zone i , based on social-economic characteristics of the zone, $\hat{Q}_i(t)$;

2. a demand elasticity factor $D_{ij}(t)$, representing the induced effect of accessibility between zones i and j on trip generation;
3. a simultaneous mode m , destination j and route r choice model, $G_{ijmr}(t)$.

Hence, part $D_{ij}(t)$ of the time-varying potential trip generation capacity $\check{Q}_i(t)$ in zone i is distributed over the different mode and route choice-options for the corridor (i, j) , using a discrete-choice model $G_{ijmr}(t)$ and converted to vehicles in [pcuh⁻¹], applying a vehicle occupancy factor θ_{1m} , as before:

$$\hat{V}_i(t) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^l} \theta_{1m} (\check{Q}_i(t) D_{ij}(t) G_{ijmr}(t)), \quad \forall l. \quad (5.32)$$

The general structure of the trip generation potential $\check{Q}_i(t)$ is similar to equation (5.23), while the simultaneous discrete-choice model $G_{ijmr}(t)$ is equal to equation (5.24), hence they are not repeated here.

The maximum amount of revealed travel demand on a corridor (part of it may actually go to another destination j , as seen in equation (5.24)), or origin-destination pair (i, j) will now be: $\check{Q}_i(t) \cdot D_{ij}(t)$, where the elasticity factor $D_{ij}(t)$ is a slight variation to equation (3.19), i.e.:

$$D_{ij}(t) = a_1 + b_1 \exp(-\lambda_2 (c_{ij}^*(t) - c_{ij}^0)), \quad \forall i, j, \quad (5.33)$$

with $c_{ij}^*(t)$ the composite costs of travel, a_1 the minimum ratio of trips (of $\check{Q}_i(t)$) that will be performed irrespective of the composite costs and $(a_1 + b_1)$ the maximum ratio, i.e. 100% of the trips that can be performed at ideal circumstances, or free-flow conditions, which is at c_{ij}^0 .

The composite cost, or logsum cost, on a origin-destination pair (i, j) are calculated by aggregating generalised costs (expressed in the utility function $u_{ijmr}(t)$, which is equation (5.26)) over all modes m and routes r that serve the origin-destination pair, analogue with equations (3.20) to (3.22), while being made time-dependent:

$$c_{ij}^*(t) = -\frac{1}{\lambda_3} \ln \left[\sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \exp(-\lambda_3 u_{ijmr}(t)) \right], \quad \forall i, j, \quad (5.34)$$

where λ_3 is the scale parameter to obtain the expected maximum utility of the choice-set: $\mathcal{C} = (\mathcal{R}_{ij}, \mathcal{M})$.

The ‘free-flow’ composite costs c_{ij}^0 can be derived by calculating the generalised costs for all choice-options under free-flow conditions, and applying equation (5.34) once.

The general solution to this model is similar to the solution of previous model, i.e.:

$$V_l(t) = e^{-\gamma_1(t-t_0)} \left(V_{l0} + \gamma_1 \int_{t_0}^{t_1} \hat{V}_l(t') e^{\gamma_1 t'} dt' \right), \quad \forall l, \quad (5.35)$$

with V_{l0} again the initial state for $V_l(t)$.

A dynamic infrastructure supply model with capacity control (B1)

Travel time on a road link is directly related to the supply of infrastructure, in other words the capacity available to the trip maker, as expressed through the volume-over-capacity (V_l/C_l^e) ratio in for example equation (1.17).

From this it is obvious that the performance of a road link in terms of the volume-over-capacity ratio can be improved by expanding the capacity of the existing road link:

$$\frac{dC_l^e(t)}{dt} = U_l^c(t), \quad \forall l, \quad (5.36)$$

where $U_l^c(t)$ is a continuous control variable $u(t)$ for *new capacity* with dimension $[\text{pcuh}^{-1} \text{T}^{-1}]^5$. Hence, $U_l^c(t)$ is changing the functionality of the road link as in the continuous network design problem (CNDP), as discussed in paragraph 4.3.3, figure 4.5.

A dynamic infrastructure supply model with capacity control and natural deterioration (B2)

Infrastructure that is in existence will loose quality at a constant rate ξ_1 in time. This natural deterioration can be due to climatological reasons as well as average known wear by vehicles on the infrastructure. The deterioration factor can be determined on the basis of the *design life* of the pavement. If the total deterioration due to these factors is linear with the existing capacity, equation (5.36) can be rewritten as:

$$\frac{dC_l^e(t)}{dt} = U_l^c(t) - \xi_1 C_l^e(t), \quad \forall l. \quad (5.37)$$

Assuming that the autonomous deterioration can be counteracted partly by conducting pavement maintenance, another control variable $u(t)$, representing the rate of maintenance $U_l^m(t)$, can be added to equation (5.37):

⁵The dimension of $U_l^c(t)$ is $[\text{pcuh}^{-1} \text{T}^{-1}]$, if T is the time-unit of integration, which may be months or years in the case of strategic modelling.

$$\frac{dC_l^e(t)}{dt} = U_l^c(t) - (\xi_1 - U_l^m(t)) C_l^e(t), \quad \forall l, \quad (5.38)$$

which is:

$$\frac{dC_l^e(t)}{dt} = U_l^t(t) - \xi_1 C_l^e(t), \quad \forall l, \quad (5.39)$$

when including the general capacity enhancement variable $U_l^t(t)$ for construction and maintenance as a control variable $u(t)$:

$$U_l^t(t) = U_l^c(t) + U_l^m(t) C_l^e(t), \quad \forall l, \quad (5.40)$$

following Lensink (2002).

The solution to this equation obviously is:

$$C_l^e(t) = e^{-\xi_1(t-t_0)} \left(C_{l_0}^e + \int_{t_0}^{t_1} U_l(t') e^{\xi_1 t'} dt' \right), \quad \forall l, \quad (5.41)$$

with $C_{l_0}^e$ the initial state for effective capacity $C_l^e(t)$.

A dynamic infrastructure supply model with capacity control and maintenance (B3)

The effective link capacity can be seen as a function of the *design capacity* $C_l^d(t)$ and pavement condition $I_l(t)$:

$$C_l^e(t) = \alpha_3 \frac{C_l^d(t)}{I_l(t)^{\beta_5}}, \quad \forall l, \quad (5.42)$$

with parameters α_3 , a scaling parameter to convert link condition units to capacity units and β_5 , the pavement condition elasticity for effective capacity. This equation is derived from Haas and Hudson (1978), who first related design capacity and pavement condition to effective capacity. Obviously, at perfect pavement condition, that is when: $I_l(t) = I_p$, or: $\alpha_3/I_p^{\beta_5} = 1.0$.

Here, the design capacity $C_l^d(t)$ is the capacity as constructed, thus assuming perfect maintenance level, and can be changed due to new construction, as before in equation (5.36):

$$\frac{dC_l^d(t)}{dt} = U_l^c(t), \quad \forall l. \quad (5.43)$$

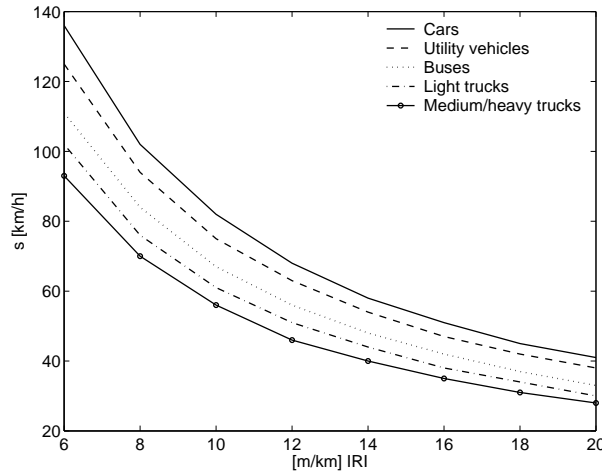


Figure 5.10: Maximum speeds as function of roughness IRI.

The functional performance of a road pavement $I_l(t)$ relates to the degree of serviceability of the pavement over time, usually indicated as ride quality or roughness. Roughness is readily noticed by road users because of its immediate impact on the riding comfort. It may also be perceived by its influence on fuel consumption.

To illustrate the effect of roughness on the performance level, figure 5.10 shows maximum speeds as function of roughness, expressed through the International Roughness Index (IRI), for a study in Brazil, as mentioned in Paterson (1987).

This International Roughness Index $I_l(t)$ is an international standard used to measure pavement roughness and is based on an open-ended scale from zero for a true planar surface, increasing to about 6 for moderately rough paved roads, 12 for extremely rough paved roads with potholing and patching, up to about 20 for extremely rough unpaved roads, as discussed in Paterson (1987).

As is also done in for example BTE (1996) a deterministic approach involving determining pavement performance by a function, which is directly relating pavement condition to variables as traffic volume, measures of pavement strength, pavement age and environmental factors, is applied here.

An empirical formula to predict the roughness progression $R_l(t)$ is then, following Paterson (1987):

$$R_l(t) = w \exp(v t)(1 + S_l)^{\beta_6} W_l(t), \quad \forall l. \quad (5.44)$$

The road roughness increases due to climate factor v , road type w , the modified structural number S_l , with factor β_6 , and through wear (in 10^6 [axlesyear⁻¹]).

The wear $W_l(t)$ on a road link is due to traffic volumes and an equivalent standard axle load factor (ESAL), Φ_{1m} , per vehicle type m , which is:

$$W_l(t) = \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^l} \Phi_{1m} \theta_{1m} (\check{Q}_i(t) D_{ij}(t) G_{ijmr}(t)), \quad \forall l. \quad (5.45)$$

Parameter α_4 converts the peak hour axle loads to an equivalent daily or yearly measure (remember that the state variables represent peak hour conditions only). The equivalent standard axle load factor per vehicle type m is exogenously derived. Parameter φ_{my} is the average load on an axle y by vehicle type m , φ_y^S is the standard single axle load of axle group y and ρ_1 the axle load equivalency exponent:

$$\Phi_{1m} = \sum_m \left(\frac{\varphi_{my}}{\varphi_y^S} \right)^{\rho_1}, \quad \forall m. \quad (5.46)$$

The modified structural number, S_l , of the road link is exogenously derived:

$$S_l = 0.40 \log_{10} T_l^D + 3.0, \quad \forall l, \quad (5.47)$$

where T_l^D is the amount of design traffic that goes with the design capacity C_l^d .

Combining these, the change in road condition $I_l(t)$ is modelled due to wear, i.e. roughness due to axle loading and climatological effects, $R_l(t)$, as:

$$\frac{dI_l(t)}{dt} = \gamma_2 R_l(t), \quad \forall l, \quad (5.48)$$

which can be controlled by applying the following control function $M_l(t)$ to the maintenance 'gap' ($I - I_p$), which is:

$$M_l(t) = \alpha_5 (I_l(t) - I_p) U_l^m(t)^{\beta_7}, \quad \forall l, \quad (5.49)$$

where I_p is the perfect pavement condition ($I_p = 2 \text{m km}^{-1}$ IRI), $U_l^m(t)$ is the control variable known as a pavement overlay and α_5 and β_7 are parameters.

Equation (5.49) can be derived from experimental results described in Colucci-Rios and Sinha (1985), who found that the relationship between rate of reduction in roughness R and resurfacing through overlay thickness U^m (in their publication in [inches]) is:

$$\Delta R = -\alpha_5 (U^m)^{\beta_7}, \quad (5.50)$$

with: $\alpha_5 = 0.61$ and: $\beta_7 = 0.26$.

Logically, the following dynamic equation can now be constructed:

$$\frac{dI_l(t)}{dt} = \gamma_2 (R_l(t) - M_l(t)), \quad \forall l. \quad (5.51)$$

Furthermore, natural deterioration using factor ξ_1 can be introduced again, similar to equation (5.37). In addition, Donaghy and Schintler (1998) suggest (for such equation) that the link maintenance level should be corrected for the relative pavement improvement (seen over a cross-section of the road link) due to capacity expansion. For better understanding both processes are introduced explicitly writing the trajectory for $I_l(t)$ over time interval (t_0, t_1) . Also, adjustment coefficients γ_2 are applied to delay the contributions of the different processes:

$$\begin{aligned} I_l(t_1) = & I_l(t_0) + \gamma_2 \int_{t'_0}^{t'_1} R_l(t') dt' + \gamma_2 \xi_1 \int_{t'_0}^{t'_1} I_l(t') dt' - \gamma_2 \int_{t'_0}^{t'_1} M_l(t') dt' \\ & - \gamma_2 \alpha_6 \int_{t'_0}^{t'_1} (I_l(t) - I_p) \frac{U_l^c(t')}{C_l^d(t')} dt', \quad \forall l, \end{aligned} \quad (5.52)$$

with weight parameter α_6 . After taking the time-derivative this yields:

$$\frac{dI_l(t)}{dt} = \gamma_2 \left(R_l(t) + \xi_1 I_l(t) - M_l(t) - \alpha_6 (I_l(t) - I_p) \frac{U_l^c(t)}{C_l^d(t)} \right), \quad \forall l. \quad (5.53)$$

System induced effects

The travel demand and infrastructure supply dynamics discussed in previous paragraphs will influence the transport network performance over time. Hence, it seems logical to introduce some dynamic equations that measure the induced effects of the disequilibrium transport dynamics.

In this model the system induced effects are related to the 'aggregated' disequilibrium level $\hat{V}_l(t)$ only. For explicitly deriving system induced effects it should, however, be possible to make a distinction in vehicle-type m and possibly also in population segmentation k . To do so, separate state variables $V_{lmk}(t)$ will have to be defined. This would, however, imply a too large increase of $m \times k$ times as many state-variables as well as costate variables. Hence, here the system induced effects are first calculated at the disaggregate mode and population segmentation level, before being accumulated into the equilibrium level $\hat{V}_l(t)$.

The change in person-throughput $P(t)$ on a route, as discussed in paragraph 3.2.2, can be modelled as:

$$\frac{dP(t)}{dt} = \gamma_1 \left(\hat{P}(t) - P(t) \right). \quad (5.54)$$

Person throughput. as discussed before in paragraph 3.2.2, in the transport network is calculated as the number of person trips made (so leaving out the vehicle occupancy factor θ_{1m}) on all routes r multiplied by the speed of movement $s_r(t)$:

$$\hat{P}(t) = \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} (\check{Q}_i(t) D_{ij}(t) G_{ijmr}(t) s_r(t)), \quad (5.55)$$

with α_4 a parameter to convert the peak-hour traffic to an equivalent daily or yearly person-throughput, and where the speed is:

$$s_r(t) = \frac{d_r}{\tau_r(t)}, \quad \forall r. \quad (5.56)$$

The dimension of $P(t)$ is in $[\text{perskmh}^{-2}]$, since $P(t)$ is a continuous variable as compared to $P_r^{\Delta t}$ in equation (3.4), which is measured over a period of time Δt , hence has dimensions $[\text{perskmh}^{-1}]$.

The total number of kilometres $K(t)$ driven in the transport network can be derived similarly as $P(t)$, that is:

$$\frac{dK(t)}{dt} = \gamma_1 (\hat{K}(t) - K(t)), \quad (5.57)$$

with the disequilibrium ‘mileage’ being:

$$\hat{K}(t) = \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \theta_{1m} (\check{Q}_i(t) D_{ij}(t) G_{ijmr}(t) d_r). \quad (5.58)$$

To calculate the total non-point emissions for pollutant $p \in \mathcal{P}$ as discussed in paragraph 3.4, another dynamic equation is used:

$$\frac{dE(t)}{dt} = \gamma_1 (\hat{E}(t) - E(t)). \quad (5.59)$$

The disequilibrium level of emissions $\hat{E}(t)$ is calculated as the number of mode specific trips multiplied with a mode and pollutant-specific emission factor $\epsilon_{m|p}$, link length d_l and a speed factor (that assumes emissions are lower at higher speeds in network), using parameter $\beta_{\text{sm}|p}$, following equation (1.22) by Zietsman (2000), which is also applied in Kim and Hoskote (1983), with parameter values given in equations (3.25) to (3.27):

$$\begin{aligned} \hat{E}(t) = & \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \sum_{l_r \in r} \theta_{1m} (\check{Q}_i(t) D_{ij}(t) G_{ijmr}(t)) \\ & \epsilon_{m|p} d_{l_r} \left(\frac{d_{l_r}}{\tau_{l_r}(t)} \right)^{-\beta_{\text{sm}|p}}, \end{aligned} \quad (5.60)$$

with link travel times as before:

$$\tau_l(t) = \tau_l^0 \left[1.0 + \alpha_1 \left(\frac{V_l(t)}{C_l^e(t)} \right)^{\beta_1} \right], \quad \forall l. \quad (5.61)$$

Equation (5.60) is formulated for use of one, possibly dominant, pollutant type only, for example introducing a composite emission factor. If one wants to include several types of pollutants an extra summation over $p \in \mathcal{P}$ is required.

5.3.2 Transport policy objectives

In an optimal control formulation the so-called equations of motion for the state variables $x(t)$, which are discussed in previous paragraph, are controlled by the control variables $u(t)$ as to minimise or maximise a certain cost criterion J . These cost functions represent (combinations of) transport policy objectives, which are accordingly formulated in terms of transport system performance or movement needs. Hence, the sustainable transport system requirements mentioned in chapter 3 can be reformulated as cost criteria in the optimal control model.

A cost criterion for controlling traffic performance (C1)

As the policy maker wants to focus on the actual movement of vehicles in transport systems, a cost criterion that is directly related to measures of congestion, or level-of-service seems appropriate.

Hence, the cost criterion could be formulated as to keep or bring the level of service for a road link or all links in the network at a certain volume-over-capacity level δ_l over the time horizon $(t_0 - t_1)$, which is, following Donaghy and Schintler (1998):

$$\begin{aligned} \min \int_{t_0}^{t_1} \sum_l \left(\frac{V_l(t)}{C_l^e(t)} - \delta_l \right)^2 dt = \\ \max - \int_{t_0}^{t_1} \sum_l \left(\frac{V_l(t)}{C_l^e(t)} - \delta_l \right)^2 dt. \end{aligned} \quad (5.62)$$

The implication of this cost criterion is that the control paths are chosen as such that all existing infrastructure is used homogeneously, that is links with a low volume-over-capacity level will increasingly be used to reach level δ_l , whereas the contrary applies to links with a high volume-over-capacity level.

A cost criterion for controlling accessibility (C2)

When a policy maker wants to focus on the interaction opportunities of people the accessibility concept by Sales Filho (1998), as discussed in paragraph 3.2.3 could be used, also including the theory on composite costs.

This accessibility cost criterion aims at minimising composite costs between the origin-destination zone pairs that have highest attraction value: $\sqrt{\tilde{Q}_i(t)\tilde{X}_j(t)}$, hence maximising over the time horizon $(t_0 - t_1)$:

$$\max \int_{t_0}^{t_1} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \left(\tilde{Q}_i(t) \tilde{X}_j(t) \right)^{\frac{1}{2}} \exp(-\beta_9 c_{ij}^*(t)) dt, \quad (5.63)$$

where β_9 is a parameter, $\tilde{Q}_i(t)$ and $\tilde{X}_j(t)$ are exogenous variables for corrected origin and destination attractiveness, as in equations (3.10) and (3.11), and $c_{ij}^*(t)$ is the composite cost of travel between zones i and j , which is equation (5.34):

$$c_{ij}^*(t) = -\frac{1}{\lambda_3} \ln \left[\sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \exp(-\lambda_3 u_{ijmr}(t)) \right], \quad \forall i, j. \quad (5.64)$$

The implication of this cost criterion obviously is that resource allocation is directed at relatively attractive origin-destination pairs (i, j) .

A cost criterion for controlling person-throughput (C3)

If the focus of transport policy is on the available network capacity to accommodate quick and comfortable movement of persons by the existing modes of transport, the person throughput over time could be aimed at, i.e. maximising person throughput in the transport network over the time horizon $(t_0 - t_1)$:

$$\max \int_{t_0}^{t_1} P(t) dt, \quad (5.65)$$

which is equal to the productive capacity of the transport network C^P in [perskmh⁻²], analogue to equation (3.7). The person-throughput $P(t)$ is calculated using equation of motion (5.54).

The implication of this cost-criterion is amongst others that modes with a high vehicle-occupancy factor θ_{1m} are given most resources.

A cost criterion for controlling equity (C4)

When equity is considered by the policy maker, the composite costs of travelling between an origin and destination for the different population segmentations k should be roughly equal, as expressed before in equation (3.16).

Hence, for $k \in \mathcal{K}$ segmentations, the equity level that should be aimed at is minimising the mean deviation:

$$\begin{aligned} & \min \int_{t_0}^{t_1} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \left(\frac{1}{K} \sum_{k \in \mathcal{K}} | \check{Q}_{i|k}(t) c_{ij|k}^*(t) - \bar{c}_{ij}^*(t) | \right) dt = \\ & \max \int_{t_0}^{t_1} - \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \left(\frac{1}{K} \sum_{k \in \mathcal{K}} | \check{Q}_{i|k}(t) c_{ij|k}^*(t) - \bar{c}_{ij}^*(t) | \right) dt, \end{aligned} \quad (5.66)$$

over the time horizon $(t_0 - t_1)$, with $\bar{c}_{ij}^*(t)$, as before, being the weighted average composite costs of travel on origin-destination pair (i, j) for all population segmentations k , which is:

$$\bar{c}_{ij}^*(t) = \frac{1}{K} \sum_{k \in \mathcal{K}} \left(\check{Q}_{i|k}(t) c_{ij|k}^*(t) \right), \quad \forall i, j, \quad (5.67)$$

while, the composite costs of equation (5.34) are now segmentation k dependent as well, that is:

$$c_{ij|k}^*(t) = -\frac{1}{\lambda_{3|k}} \ln \left[\sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \exp(-\lambda_{3|k} u_{ijmr|k}(t)) \right], \quad \forall i, j, k. \quad (5.68)$$

and, the production potential share for segment k :

$$\check{Q}_{i|k}(t) = \frac{\check{Q}_{i|k}(t)}{\sum_{k' \in \mathcal{K}} \check{Q}_{i|k'}(t)}. \quad (5.69)$$

However, for computational reasons, which are discussed later on, equation (5.66) is better written in terms of quadratic deviation from the average segmentation specific composite costs at $t = t_0$, which is:

$$\max \int_{t_0}^{t_1} - \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \left(\sum_{k \in \mathcal{K}} \check{Q}_{i|k}(t) c_{ij|k}^*(t) - \bar{c}_{ij}^*(t_0) \right)^2 dt. \quad (5.70)$$

The replacement of $\bar{c}_{ij}^*(t)$ with $\bar{c}_{ij}^*(t_0)$ is done to obtain a more stable objective function and basically implies that deterioration from the average initial level of equity is not allowed.

The general implication of this cost criterion is that preference is given to segmentations k that experience relatively high composite cost, though irrespective of the interaction opportunities at the origin and destination zones.

As mentioned before, a more equitable or levelled distribution of transport opportunities to the different segmentations results from this policy objective.

The weighted average composite costs acting as a source-term, could however also imply that investments are only directed at levelling the absolute deviation towards zero only. An overall improvement of the transport system somehow is not reflected in this measure; only the distribution of transport opportunities.

Note that to calculate this cost criterion, obviously, also the demand model needs to be changed, in order to allow for k different population segments, which is:

$$\hat{V}_l(t) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^l} \sum_{k \in \mathcal{K}} \theta_{1m} (\tilde{Q}_{i|k}(t) D_{ij|k}(t) G_{ijmr|k}(t)), \quad \forall l. \quad (5.71)$$

5.3.3 Measures and bounds

In the previous paragraphs several policy control measures $u(t)$ have been used. For each of the control variables it is common, even necessary in the case of linear controls, to pose inequality bounds on their values.

For the new construction control variable $U_l^c(t)$ a lower and upper bound (say land-use limitation, i.e. space available for expanding capacity) can be formulated as follows:

$$0 \leq U_l^c(t) \leq U_l^{c\max}, \quad \forall l. \quad (5.72)$$

Likewise the pavement overlay control $U_l^m(t)$ is bounded by a lower and upper bound (say technical limitation).

$$0 \leq U_l^m(t) \leq U_l^{m\max}, \quad \forall l. \quad (5.73)$$

The upper bound $U_l^{m\max}$, in the case of model (B3), varies with time as it is not possible to improve IRI beyond perfect pavement condition I_p , which is:

$$I_l(t) - I_p + \gamma_2 \left(R_l(t) + \xi_1 I_l(t) - \alpha_5 (I_l(t) - I_p) U_l^m(t)^{\beta_7} - \alpha_6 (I_l(t) - I_p) \frac{U_l^c(t)}{C_l^d(t)} \right) = 0, \quad \forall l, \quad (5.74)$$

or:

$$U_l^{m\max} = \left(\frac{I_l(t) - I_p + \gamma_2 \left(R_l(t) + \xi_1 I_l(t) - \alpha_6 (I_l(t) - I_p) \frac{U_l^c(t)}{C_l^d(t)} \right)}{\gamma_2 \alpha_5 (I_l(t) - I_p)} \right)^{\frac{1}{\beta_7}}, \quad \forall l. \quad (5.75)$$

Both controls $U_l^c(t)$ and $U_l^m(t)$ may also be bounded by an exogenously determined budget constraint $F^*(t)$. If the standard price for construction cost is c_1^c money units and for maintenance cost c_2^m money units, this is:

$$\sum_l (c_1^c d_l U_l^c(t) + c_2^m d_l C_l^d(t) U_l^m(t)) \leq F^*(t). \quad (5.76)$$

This budget constraint may also be formulated as an isoperimetric or integral equality or inequality constraint (see also paragraph 5.4.1):

$$\int_{t_0}^{t_1} \sum_l (c_1^c d_l U_l^c(t) + c_2^m d_l C_l^d(t) U_l^m(t)) dt = \tilde{F}_{t_1}^*, \quad (5.77)$$

which can then be written as:

$$\frac{d\tilde{F}(t)}{dt} = \sum_l (c_1^c d_l U_l^c(t) + c_2^m d_l C_l^d(t) U_l^m(t)), \quad (5.78)$$

with $\tilde{F}(t)$ being the integral expenditure. The boundary points for $\tilde{F}(t)$ are:

$$\tilde{F}(t_0) = 0, \quad \tilde{F}(t_1) \leq \tilde{F}_{t_1}^*. \quad (5.79)$$

Both linear tax controls $U^v(t)$ and $U_j^p(t)$ should be bounded by inequality constraints as well:

$$0 \leq U^v(t) \leq U^{v\max}, \quad (5.80)$$

and:

$$0 \leq U_j^p(t) \leq U_j^{p\max}, \quad \forall j. \quad (5.81)$$

Furthermore, pure state constraint may be applied to the emissions $E(t)$ as well. The environmental capacity E^* is then the exogenously determined maximum total emissions at any time t :

$$E(t) \leq E^*. \quad (5.82)$$

Likewise, the emission constraint can be written as an integral inequality constraint:

$$\int_{t_0}^{t_1} E(t) dt \leq \tilde{E}_{t_1}^*, \quad (5.83)$$

which can then be written as:

$$\frac{d\tilde{E}(t)}{dt} = E(t), \quad (5.84)$$

with $\tilde{E}(t)$ being the integral emissions. The boundary points for $\tilde{E}(t)$ are:

$$\tilde{E}(t_0) = 0, \quad \tilde{E}(t_1) \leq \tilde{E}_{t_1}^*. \quad (5.85)$$

Alternatively, a pure state constraint at the *endpoint* may be applied to the emissions $E(t_1)$. The environmental capacity $E_{t_1}^*$ is then the exogenously determined maximum total emissions at time $t = t_1$:

$$E(t_1) \leq E_{t_1}^*. \quad (5.86)$$

Summarising, a model has been described consisting of travel demand dynamics and infrastructure supply dynamics, where infrastructure is controlled through construction and maintenance and demand through vehicle taxes and parking taxes. In the meantime some system effects (as vehicle kilometres, person throughput and total emissions) are monitored. The controls are selected as such as to maximise (or minimise) the transport objective function. These controls are constrained by physical, financial or technical bounds. The emissions as well as budget are bound over time or at the end of the planning horizon t_1 and should not exceed a certain value, while at the same time the model is trying to optimise the objective. Hence, a dynamic optimisation model is described that can be solved for example using Pontryagin's Maximum Principle.

5.4 Pontryagin's Maximum Principle

Optimal control was originally developed in the Soviet Union by Lev Pontryagin and others in the early 1960s (Pontryagin et al., 1962) and has been applied in both technical sciences (often as minimisation problems, like fuel control for soft landing of a rocket on the moon) as well as in social, medical and economic sciences (for example maximising the effectiveness of drugs treatment). A basic set of necessary conditions for optimal control is named *Pontryagin Maximum Principle*, shortly Maximum Principle, or in the case of minimisation problems, Minimum Principle.

Besides the Maximum principle, another popular approach to optimal control is called dynamic programming as developed by Richard Bellman, which is based on the principle that given the fact that an optimal path has the property that whatever the initial conditions and control values are (over some initial period), the control over the remaining period must be optimal for the remaining

problem, with the state variables resulting from earlier decisions, considered as the initial condition for the new decisions. Dynamic programming has been used extensively in discrete-time problems. The method has as an advantage that it can better deal with non-smooth functions. The greatest disadvantage, however, which is determinative for not choosing this method in this research, is the so-called curse of dimensionality. That is, dynamic programming solves for the optimal solution from every feasible state, which is the solution of the value function (the minimal value of the cost criterion from some time to the final time), implying a large number of feasible states. And a large number of feasible states is also the case here, which means a very long time is required to solve a problem.

Pontryagin's Maximum Principle⁶ can be used to solve the general constrained optimal control problem. Assume that there are r control variables, n state variables, and m constraints on the control variables. Suppose that the first m' constraints are inequality constraints and the remaining $m - m'$ constraints are equality constraints. Generally written the optimal control problem involving continuous-time, a finite time horizon $t \in [t_0, t_1]$, a vector of state variables $\vec{x}(t)$, a vector of control variables $\vec{u}(t)$, inequality constraints $g^j(t)$, equality constraints $g^k(t)$ and a cost criterion J (assuming the final state: $F(x(t_1), t) = 0$ in equation (5.3)) to be maximised:

$$\begin{aligned} \max \quad & \int_{t_0}^{t_1} J(\vec{x}(t), \vec{u}(t), t) dt \\ \text{s.t.} \quad & \frac{dx_i(t)}{dt} = f^i(\vec{x}(t), \vec{u}(t), t), \quad i = 1, 2, \dots, n, \\ & g^j(\vec{x}(t), \vec{u}(t), t) \geq 0, \quad j = 1, 2, \dots, m', \\ & g^k(\vec{x}(t), \vec{u}(t), t) = 0, \quad k = m' + 1, \dots, m, \\ & x_i(t_0) = x_{it_0}, \quad x_i(t_1) = x_{it_1}, \quad i = 1, 2, \dots, n. \end{aligned} \tag{5.87}$$

The time horizon t_1 and the initial and final values of the state variables are exogenously specified.

The set of admissible controls to this system is:

$$\begin{aligned} \mathcal{W}(\vec{x}^*(t), t) \equiv \{ \vec{u}(t) | & g^j(\vec{x}^*(t), \vec{u}(t), t) \geq 0, \quad j = 1, 2, \dots, m'; \\ & g^k(\vec{x}^*(t), \vec{u}(t), t) = 0, \quad k = m' + 1, \dots, m \}, \end{aligned} \tag{5.88}$$

with $\vec{x}^*(t)$ the optimal path for the state variables.

Above mentioned type of constraints is usually called mixed constraints as they involve both state variables as well as control variables. Constraints are called pure state constraints when they are only dependent on state variables, or:

⁶The text in this paragraph is based on several descriptions of Pontryagin's Maximum Principle, which can be found in standard text-books on optimal control, for example Léonard and Van Long (1992).

$$g^l(\vec{x}(t), t) \geq 0, \quad l = m + 1, \dots, m''. \quad (5.89)$$

In general, these constraints should satisfy the constraint qualification. The most convenient constraint qualification is the *rank condition*, which requires that the number of active constraints q not be greater than the number of control variables p . Or more formally stated, the following full-rank condition:

$$\text{rank}\left[\frac{\partial g^j(\vec{x}^*(t), \vec{u}^*(t), t)}{\partial u_p(t)}, \text{diag}(g)\right] = q \leq p, \quad j = 1, \dots, m', \quad (5.90)$$

holds for all arguments $(x_i(t), u_p(t), t)$. This condition means that the gradients with respect to the control value $u_p(t)$ of all active constraints in problem (5.87) must be linearly independent.

Note that the general optimal control problem (5.87), can be characterised as *non-autonomous* as it does depend on time t explicitly. Every non-autonomous system, however, can be converted to an autonomous system by adding one dimension for time t . In other words, if the equation of motion is:

$$\frac{dx_i(t)}{dt} = f^i(\vec{x}(t), \vec{u}(t), t), \quad i = 1, 2, \dots, n, \quad (5.91)$$

it can be written as an autonomous system with $n + 1$ state variables by substituting: $x_{n+1} \equiv t$, that is:

$$\frac{dx_{n+1}(t)}{dt} = 1, \quad (5.92)$$

with boundary point:

$$x_{n+1}(t_0) = t_0. \quad (5.93)$$

The mathematical problem that remains is generally solved applying a *Lagrangean* equation \mathcal{L} , also called augmented Hamiltonian:

$$\mathcal{L} = J(\vec{x}(t), \vec{u}(t), t) + \sum_i^n \mu_i(t) f^i(\vec{x}(t), \vec{u}(t), t) + \sum_j^m \omega_j(t) (g^j(\vec{x}(t), \vec{u}(t), t)), \quad (5.94)$$

where the Hamiltonian is defined as:

$$\mathcal{H}(\vec{x}(t), \vec{u}(t), t) \equiv J(\vec{x}(t), \vec{u}(t), t) + \sum_i^n \mu_i(t) f^i(\vec{x}(t), \vec{u}(t), t). \quad (5.95)$$

The state variables in the Hamiltonian are associated with an auxiliary variable called a costate variable denoted by $\mu_i(t)$. Similarly a Lagrange multiplier $\omega_j(t)$ is associated with each constraint $g^j(t)$ in equation (5.94). The value of these costate variables and multipliers is often interpreted as the shadow price of an associated state variable or constraint.

The Lagrangean \mathcal{L} in equation (5.94) is maximised with respect to $u_p(t)$ subject to the (in)equality constraints. Most importantly for maximising the Lagrangean, the *Kuhn-Tucker* optimality conditions need to apply, that is:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \omega_j(t)} &= g^j(t) \geq 0, \quad j = 1, 2, \dots, m', \\ \omega_j(t) &\geq 0, \quad j = 1, 2, \dots, m', \\ \omega_j(t) \frac{\partial \mathcal{L}}{\partial \omega_j(t)} &= 0, \quad j = 1, 2, \dots, m',\end{aligned}\tag{5.96}$$

which are also known as the complementary-slackness (CS) conditions in static optimisation, and ensure that whenever the j th Lagrange multiplier is nonzero, the j th constraint is satisfied as a strict equality, and whenever the j th constraint is a strict inequality, the j th Lagrange multiplier is zero. For convex functions and convex constraints, these conditions can be shown to be necessary and sufficient for a global optimum solution.

These conditions imply in practice that the transport planner will not use a stock $x_i(t)$ to its maximum if the marginal price or contribution to the functional turn zero.

The Maximum Principle provides a set of necessary conditions for optimality and defines a class of problems for which the Maximum Principle is sufficient for optimality. The necessary conditions provide a set of candidates for optimality, whereas the sufficiency conditions guarantee that a candidate satisfying these sufficiency conditions is optimal. The necessary and sufficient conditions are adapted from Léonard and Van Long (1992):

Necessity Let $\bar{u}^*(t)$ be an optimal solution to the constrained problem (5.87) and $\bar{x}^*(t)$ be the corresponding optimal time path of the state variables. Then there exist costate variables $\bar{\mu}(t)$ and (assuming the rank condition (5.90) is satisfied) multipliers $\bar{\omega}(t)$ such that:

1. At any time t for given $x_i^*(t)$ and $\mu_i(t)$, the control variables $u_p^*(t)$ maximise the Hamiltonian (5.95) subject to the condition that $u_p(t)$ belong to the set of admissible controls $\mathcal{W}(\bar{x}(t), t)$ in equation (5.88). In view of the rank condition this implies that there exist multipliers $\bar{\omega}(t)$ such that:

$$\frac{\partial \mathcal{L}^*}{\partial u_p(t)} = 0, \quad p = 1, 2, \dots, r,\tag{5.97}$$

$$\begin{aligned} \omega_j(t) &\geq 0, \quad g^j(\bar{x}^*(t), \bar{u}^*(t), t) \geq 0, \quad j = 1, 2, \dots, m', \\ \omega_j(t)g^j(\bar{x}^*(t), \bar{u}^*(t), t) &= 0, \quad j = 1, 2, \dots, m', \end{aligned} \quad (5.98)$$

$$g^k(\bar{x}^*(t), \bar{u}^*(t), t) = 0, \quad k = m' + 1, \dots, m, \quad (5.99)$$

where the asterisk on \mathcal{L} indicates that the derivatives are evaluated at $(\bar{x}^*(t), \bar{u}^*(t))$. The multipliers $\bar{\omega}(t)$ are piecewise-continuous and continuous on each point of continuity of $\bar{u}^*(t)$, implying only a finite number of discontinuous jumps.

In some special cases \mathcal{L} is linear in the control value $u_p(t)$. If the control has a lower and upper bound, i.e.: $u_p^- \leq u_p \leq u_p^+$, expressed through the inequality constraint: $g^j(\bar{x}^*(t), \bar{u}^*(t), t) \geq 0$, then the optimal control has the form:

$$u_p^*(t) = \begin{cases} u_p^+ & \text{if } \partial \mathcal{H} / \partial u_p > 0, \\ u_p^- \leq u_p \leq u_p^+ & \text{if } \partial \mathcal{H} / \partial u_p = 0, \\ u_p^- & \text{if } \partial \mathcal{H} / \partial u_p < 0, \end{cases} \quad (5.100)$$

as:

$$\frac{\partial \mathcal{L}}{\partial u_p} = \frac{\partial \mathcal{H}}{\partial u_p} + \omega^- - \omega^+ = 0, \quad (5.101)$$

where ω^- and ω^+ are the multipliers with the lower and upper bound on u_p respectively.

The optimal control is either *bang-bang*, implying that the control variables u_p take on their extreme values only but may switch between them, or *singular*, implying that the switching function (5.100) is identically equal to zero, hence the control is indeterminate and modified methods need to be used to obtain the singular control value u_p . In the bang-bang control policy, the points where the control switches are known as switching times.

2. The costate variables $\mu_i(t)$, $i = 1, 2, \dots, n$ are continuous and have piecewise-continuous derivatives satisfying:

$$\frac{d\mu_i(t)}{dt} = -\frac{\partial \mathcal{L}^*}{\partial x_i(t)}, \quad i = 1, 2, \dots, n; \quad (5.102)$$

3. Equally, the equations of motion for $x_i(t)$ are:

$$\frac{dx_i(t)}{dt} = \frac{\partial \mathcal{L}^*}{\partial \mu_i(t)} = f^i(\bar{x}(t), \bar{u}(t), t), \quad i = 1, 2, \dots, n; \quad (5.103)$$

4. The Lagrangean $\mathcal{L}(\bar{x}^*(t), \bar{u}^*(t), \bar{\mu}(t), \bar{\omega}(t), t) \equiv \phi(t)$ is a continuous function of t . On each interval of continuity of $\bar{u}^*(t)$, $\phi(t)$ is differentiable and:

$$\frac{d\phi(t)}{dt} = \frac{d\mathcal{L}^*}{dt} = \frac{\partial \mathcal{L}^*(t)}{\partial t}; \quad (5.104)$$

5. The boundary conditions:

$$x_i(t_0) = x_{it_0}, \quad x_i(t_1) = x_{it_1}, \quad (5.105)$$

must be satisfied⁷.

Sufficiency Let $(\bar{x}^*(t), \bar{u}^*(t))$ satisfy the necessity conditions stated above and assume the Lagrangean (5.94) is concave in $(\bar{x}(t), \bar{u}(t))$, then $(\bar{x}^*(t), \bar{u}^*(t))$ is an optimal path for the problem (5.87). If \mathcal{L} is strictly concave, then $(\bar{x}^*(t), \bar{u}^*(t))$ is the unique optimal solution.

Furthermore, the concavity of the Lagrangean is assured if the following conditions are met (Léonard and Van Long, 1992):

1. Cost criterion J is concave in $(\bar{x}(t), \bar{u}(t))$;
2. Each term $\mu_i^*(t)f^i(\bar{x}(t), \bar{u}(t), t)$ is concave in $(\bar{x}(t), \bar{u}(t))$;
3. Each of the m' inequality constraints: $g^j(\bar{x}(t), \bar{u}(t), t) \geq 0$, is concave in $(\bar{x}(t), \bar{u}(t))$;
4. Each of the $m - m'$ equality constraints: $g^k(\bar{x}(t), \bar{u}(t), t) = 0$, has the property that $\omega_k^*(t)g^k(\bar{x}(t), \bar{u}(t), t)$ is concave in $(\bar{x}(t), \bar{u}(t))$.

Concavity of a function can be checked through the *Hessian matrix* of second-order derivatives. The Hessian matrix should be negative-semidefinite everywhere. For a minimisation problem these conditions are turned around, implying convexity requirements as well as positive-semidefiniteness. See for a discussion on convexity and concavity of functions as well as further sufficiency conditions, appendix C. Furthermore, one should realise that a minimisation problem in system (5.87) would just mean a multiplication of the functional J by minus one, as for example seen in equation (5.62).

5.4.1 Transversality conditions and path constraints

In system (5.87) two boundary conditions at t_0 as well as at t_1 were given, in a so-called fixed-endpoint problem, allowing the shadow price or costate value to be free at the boundary. In the type of problem studied in this research not all state variables should have a fixed boundary point. Hence, the boundary conditions should be modified to allow for some free endpoints. Hence, transversality

⁷Obviously, these are very restrictive assumptions. Hence, in paragraph 5.4.1, additional necessary conditions, or transversality conditions, are introduced, to replace one boundary condition per state variable $x_i(t)$.

conditions have to be imposed to replace one of the boundary conditions with the state variable.

Generally stated, if: $x_i(t_1) = \text{free}$, for some $i = 1, 2, \dots, n'$, hence: $x_j(t_1) = x_{jt_1}$, for $j = n' + 1, \dots, n$, the transversality conditions are⁸:

$$\mu_i^*(t_1) = 0, \quad i = 1, 2, \dots, n'. \quad (5.106)$$

This transversality condition implies that the transport planner attaches no value to the terminal stock and is not constrained to meet a certain target $x_i(t_1)$, hence the stock should be used until its marginal contribution is zero at the end of the planning horizon.

In another case the final value of some state variable may be constrained to be not less or more than a prespecified constant, for example in the case of an environmental capacity. Hence, in case of a lower bound: $x_i(t_1) \geq x_{iL}$, for some $i = 1, 2, \dots, n'$. In this case the transversality conditions for a lower bound constraint on the endpoint are, alike the Kuhn-Tucker conditions in equation (5.96):

$$\begin{aligned} \mu_i^*(t_1) &\geq 0, \quad i = 1, 2, \dots, n', \\ (x_i(t_1) - x_{iL}) &\geq 0, \quad i = 1, 2, \dots, n', \\ \mu_i^*(t_1)(x_i(t_1) - x_{iL}) &= 0, \quad i = 1, 2, \dots, n'. \end{aligned} \quad (5.107)$$

In terms of necessity and sufficiency conditions the transversality conditions should be added to these conditions. Furthermore, should: $\mu_i^*(t_1)(x_i(t_1) - x_{iL})$, be concave in the final endpoints. The derivation of these transversality conditions and sufficiency conditions is well described in standard books on optimal control, e.g. Bryson and Ho (1975), Léonard and Van Long (1992) or Kamien and Schwartz (1993).

Apart from constraints at the endpoint, the transversality conditions, constraints can also apply to intermediate points in time or over the whole time path: $t_0 \leq t \leq t_1$, rather than just at the end points.

A restriction on the overall path of the variables can be imposed using a isoperimetric or integral equality constraint, as:

$$\int_{t_0}^{t_1} G(\vec{x}(t), \vec{u}(t), t) dt = B(t_1). \quad (5.108)$$

This is basically an equality constraint that can be replaced by a differential equation and two additional boundary constraints, as follows:

⁸Formally, if a final state $F(x_i(t_1), t_1)$, as in equation (5.3), is considered: $\mu_i^*(t_1) = \frac{\partial F(x_i, t)}{\partial x} |_{t=t_1}$, $i = 1, 2, \dots, n'$, which equals equation (5.106) for system (5.87).

$$\begin{aligned} \frac{dy_i(t)}{dt} &= G(\vec{x}(t), \vec{u}(t), t), \\ y(t_0) &= 0, \quad y(t_1) = B(t_1), \end{aligned} \tag{5.109}$$

which accordingly can be treated as a standard equation of motion, with an adjoined costate variable or Lagrange multiplier. The problem has thus changed from a free endpoint to a fixed endpoint problem with respect to total emissions.

Constraints to functions of state variables and control variables can be treated equally as stated above in equation (5.107), using Kuhn-Tucker conditions. More difficulties appear if constraints are imposed on (functions of) state variables that don't have explicit dependence on the control variable(s) in the problem. These constraints are called equality or inequality pure state-space constraints. Seierstad and Sydsæter (1987) or Bryson and Ho (1975) discuss this type of constraints in depth.

5.4.2 General computational aspects

The presented theory on optimal control and its role in engineering and economics are unquestionable. However, practical applications, apart from modelling batch processes in chemical engineering, flight trajectory planning in aerospace engineering and obviously in robotics, are still scarce. The main reason for this being the level of mathematical sophistication, doubtful viability of optimisation under uncertain conditions, estimation problems and high computational requirements as stated in Schwartz (1996). However, recent developments in analytical methods (like symbolic differentiation applied here using MAPLE), accompanying computational codes (like MATLAB procedures widely available through the internet) and improved computer performance, mark a recent increase in applied optimal control studies.

The necessary and sufficient conditions for optimality as discussed before, provide the inputs to a so-called *indirect* solution approach (by directly solving these conditions), instead of directly solving the objective function and accompanying constraints as in a *direct* approach. The basic idea of the latter approach is to transform the optimal control problem into a finite dimensional optimisation problem (also known as control parametrisation), which can then be approximated as a constrained nonlinear programming problem. In some cases both type of methods are used in conjunction, where the estimates for the costate trajectories are obtained in a direct approach, which is then refined applying the indirect approach. As, only the very simple optimal control problems can be solved analytically, numerical methods should be deployed. Optimal control problems as formulated through the indirect Maximum Principle approach are often at least two-point boundary value problems, with the initial-value for the state variables (e.g. system (5.87)) as well as the boundary

values for the costate variables (e.g. equation (5.106)) known. In some cases multipoint boundary value problems appear, for example when there are pure state variable inequality constraints. Finding solutions to these nonlinear two-point or multipoint boundary value problems is in many cases not a trivial exercise. Hence, in chapter 6 different algorithms for the numerical solution are considered.

The numerical problem can be summarised as to find:

1. the n state variables, $x_i(t)$, $i = 1, 2, \dots, n$;
2. the n costate variables, $\mu_i(t)$, $i = 1, 2, \dots, n$;
3. the r control variables, $u_p(t)$, $p = 1, 2, \dots, r$;
4. the m Lagrange multipliers, $\omega_j(t)$, $j = 1, 2, \dots, m', \dots, m$.

to satisfy simultaneously:

1. the n system differential equations (involving $\vec{x}(t), \vec{u}(t)$);
2. the n costate differential equations (involving $\vec{\mu}(t), \vec{x}(t), \vec{u}(t)$);
3. the r optimality conditions (involving $\vec{\omega}(t), \vec{\mu}(t), \vec{x}(t), \vec{u}(t)$);
4. the initial and (final) boundary conditions (involving $\vec{\mu}(t), \vec{x}(t)$);
5. the h exogenous variables, $z_g(t)$, $g = 1, 2, \dots, h$,

which adds up to solving, from a given set of known parameter values, simultaneously, $2n$ first-order ordinary differential equations and $3r$ zero-order equations (assuming an upper and lower bound per control variable), while satisfying, at least, $2n$ boundary points and the complete time paths of the h exogenous variables over the relevant number of time periods.

The type of controller based on Pontryagin's Maximum Principle can be characterised as an open-loop controller (that is control without feedback). The open-loop optimal controller is depicted for the shooting method (see paragraph 6.2) in figure 5.11 (next page). In contrast there is the closed-loop optimal controller, as is basically schematized in figure 5.1, where optimal control $u^*(t)$ is stated in terms of the current state $x^*(t)$ directly. The closed-loop formulation has advantages over the open-loop formulation (mainly in terms of progression of uncertainties over the prediction interval), but is much more difficult to derive at. Given the complexity of the problem in this research an open-loop strategy is used. Several tolerance levels for the convergence of the boundary point solution as well as for the solution of the nonlinear model need to be specified. For example the boundary point for the transversality condition: $\mu(t_1) = 0$, in equation (5.106), should computationally be read as: $\mu(t_1) \leq \varepsilon$, where ε is the tolerance level for the boundary point solution. In practice this often means that the optimal costate trajectory is derived by systematically changing the

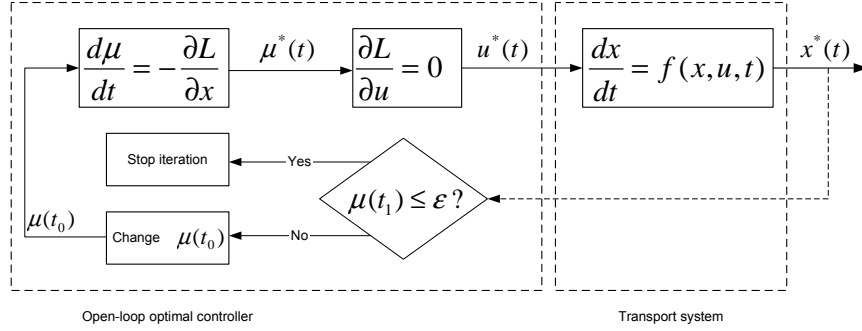


Figure 5.11: Open-loop optimal control (shooting method).

initial value $\mu(t_0)$ until the transversality condition: $\mu(t_1) \leq \varepsilon$, with: $\varepsilon \approx 0$ chosen very small, is met. Some more details on the applied computational procedures for deriving the optimal control are discussed in paragraph 6.2.

5.4.3 Deriving the optimal control (A4/B3/C1)

For deriving the Pontryagin optimality conditions the Lagrangean function is constructed here, based on the theory presented in paragraph 5.4, for one of the dynamic systems. This dynamic system, which is the most elaborated model, consists of the disequilibrium network elastic model (A4), the dynamic infrastructure supply model with capacity control and maintenance (B3), in combination with the traffic performance cost criterion (C1), the system induced effects and several bounds on the controls as well as the emission state-constraint:

$$\begin{aligned}
 & \max - \int_{t_0}^{t_1} \sum_l \left(\frac{V_l(t)}{C_l^e(t)} - \delta_l \right)^2 dt \\
 & \text{s.t. } \frac{dV_l(t)}{dt} = \gamma_1 \left(\hat{V}_l(t) - V_l(t) \right), \quad \forall l, \\
 & \quad \frac{dC_l^d(t)}{dt} = U_l^c(t), \quad \forall l, \\
 & \quad \frac{dI_l(t)}{dt} = \gamma_2 \left(R_l(t) + \xi_1 I_l(t) - M_l(t) - \alpha_6 (I_l(t) - I_p) \frac{U_l^c(t)}{C_l^d(t)} \right), \quad \forall l, \\
 & \quad \frac{dP(t)}{dt} = \gamma_1 (\hat{P}(t) - P(t)) \\
 & \quad \frac{dK(t)}{dt} = \gamma_1 (\hat{K}(t) - K(t)) \\
 & \quad \frac{dE(t)}{dt} = \gamma_1 (\hat{E}(t) - E(t)) \\
 & \quad \dots
 \end{aligned}$$

$$\begin{aligned}
& \dots \\
& \text{and } 0 \leq U_l^c(t) \leq U_l^{c\max}, \quad \forall l, \\
& 0 \leq U_l^m(t) \leq U_l^{m\max}, \quad \forall l, \\
& 0 \leq U^v(t) \leq U^{v\max} \\
& 0 \leq U_j^p(t) \leq U_j^{p\max}, \quad \forall j, \\
& \sum_l (c_1^c d_l U_l^c(t) + c_2^m d_l C_l^d(t) U_l^m(t)) \leq F^*(t), \\
& E(t_1) \leq E_{t_1}^*, \\
& \text{boundary conditions at } t_0 \text{ and } t_1,
\end{aligned} \tag{5.110}$$

The Lagrangean equation to this problem⁹ can be formulated as:

$$\begin{aligned}
\mathcal{H} = & - \sum_l \left(\frac{V_l(t)}{C_l^e(t)} - \delta_l \right)^2 + \sum_l \mu_{1l}(t) \gamma_1 (\hat{V}_l(t) - V_l(t)) + \sum_l \mu_{2l}(t) U_l^c(t) + \\
& \sum_l \mu_{3l}(t) \left(\gamma_2 (R_l(t) + \xi_1 I_l(t) - M_l(t) - \alpha_6 (I_l(t) - I_p) \frac{U_l^c(t)}{C_l^d(t)}) \right) + \\
& \mu_4(t) \gamma_1 (\hat{P}(t) - P_t) + \mu_5(t) \gamma_1 (\hat{K}(t) - K(t)) + \mu_6(t) \gamma_1 (\hat{E}(t) - E(t)).
\end{aligned} \tag{5.111}$$

The Lagrangean or augmented Hamiltonian to this problem then equals:

$$\begin{aligned}
\mathcal{L} = & \mathcal{H} + \sum_l \omega_{1l}(t) U_l^c(t) + \sum_l \omega_{2l}(t) (U_l^{c\max} - U_l^c(t)) + \sum_l \omega_{3l}(t) U_l^m(t) + \\
& \sum_l \omega_{4l}(t) (U_l^{m\max} - U_l^m(t)) + \omega_5(t) U^v(t) + \omega_6(t) (U^{v\max} - U^v(t)) + \\
& \sum_j \omega_{7j}(t) U_j^p(t) + \sum_j \omega_{8j}(t) (U_j^{p\max} - U_j^p(t)) + \\
& \eta_1(t) (F^*(t) - \sum_l (c_1^c d_l U_l^c(t) + c_2^m d_l C_l^d(t) U_l^m(t))) + \nu_{1t_1} (E_{t_1}^* - E(t_1)).
\end{aligned} \tag{5.112}$$

The first-order optimality conditions are specified by the equations of motion for $x_i(t)$, as in equation (5.103):

$$\frac{dx_i(t)}{dt} = \frac{\partial \mathcal{L}}{\partial \mu_i(t)} = f^i(\vec{x}(t), \vec{u}(t)), \quad \forall i \tag{5.113}$$

that is:

⁹Note that there is no fundamental difference between minimisation and maximisation, since: $\min_x f(x) = -\max_x (-f(x))$. For the argument x (the control vector): $\min_x f(x) = \max_x (-f(x))$ also holds, since: $x = \arg[-\max_{x'} (-f(x'))] = \arg[\max_{x'} (-f(x'))]$.

$$\frac{dV_l(t)}{dt} = \gamma_1(\hat{V}_l(t) - V_l(t)), \quad \forall l, \quad (5.114)$$

$$\frac{dC_l^d(t)}{dt} = U_l^c(t), \quad \forall l, \quad (5.115)$$

$$\frac{dI_l(t)}{dt} = \gamma_2(R_l(t) + \xi_1 I_l(t) - M_l(t) - \alpha_6(I_l(t) - I_p) \frac{U_l^c(t)}{C_l^d(t)}), \quad \forall l, \quad (5.116)$$

$$\frac{dP(t)}{dt} = \gamma_1(\hat{P}(t) - P(t)), \quad (5.117)$$

$$\frac{dK(t)}{dt} = \gamma_1(\hat{K}(t) - K(t)), \quad (5.118)$$

$$\frac{dE(t)}{dt} = \gamma_1(\hat{E}(t) - E(t)), \quad (5.119)$$

while the costate variables $\mu_i(t)$ satisfy, as in equation (5.102):

$$\frac{d\mu_i(t)}{dt} = -\frac{\partial \mathcal{L}}{\partial x_i(t)}, \quad \forall i, \quad (5.120)$$

that is:

$$\begin{aligned} \partial \mathcal{L} / \partial V_l(t) &= -d\mu_{1l}(t)/dt = \\ &= -\left(\frac{2}{C_l^e(t)}\right)\left(\frac{V_l(t)}{C_l^e(t)} - \delta_l\right) + \sum_{k \in \mathcal{L}} \mu_{1k}(t) \gamma_1 \left(\frac{\partial \hat{V}_k(t)}{\partial V_l(t)} - \frac{\partial V_k(t)}{\partial V_l(t)}\right) \\ &+ \sum_{k \in \mathcal{L}} \mu_{3k}(t) \gamma_2 \frac{\partial R_k(t)}{\partial V_l(t)} + \mu_4(t) \gamma_1 \frac{\partial \hat{P}(t)}{\partial V_l(t)} + \mu_5(t) \gamma_1 \frac{\partial \hat{K}(t)}{\partial V_l(t)} \\ &+ \mu_6(t) \gamma_1 \frac{\partial \hat{E}(t)}{\partial V_l(t)} + \sum_{k \in \mathcal{L}} \omega_{4k}(t) \frac{\partial U_l^m(t)^{\max}}{\partial V_l(t)}, \quad \forall l, \end{aligned} \quad (5.121)$$

$$\begin{aligned}
\partial\mathcal{L}/\partial C_l^d(t) &= -d\mu_{2l}(t)/dt = \\
&\frac{2}{C_l^e(t)}\left(\frac{V_l(t)}{C_l^e(t)} - \delta_l\right)\frac{V_l(t)}{C_l^d(t)} + \sum_{k \in \mathcal{L}} \mu_{1k}(t)\gamma_1 \frac{\partial \hat{V}_k(t)}{\partial C_l^d(t)} \\
&+ \sum_{k \in \mathcal{L}} \mu_{3k}(t)\gamma_2 \left(\frac{\partial R_k(t)}{\partial C_l^d(t)}\right) + \mu_{3l}(t)\gamma_2\alpha_6(I_l(t) - I_p)\frac{U_l^c(t)}{C_l^d(t)^2} \\
&+ \mu_4(t)\gamma_1 \frac{\partial \hat{P}(t)}{\partial C_l^d(t)} + \mu_5(t)\gamma_1 \frac{\partial \hat{K}(t)}{\partial C_l^d(t)} + \mu_6(t)\gamma_1 \frac{\partial \hat{E}_t}{\partial C_l^d(t)} \\
&+ \sum_{k \in \mathcal{L}} \omega_{4k}(t) \frac{\partial U_l^m(t)^{\max}}{\partial C_l^d(t)} - \eta_1(t)c_2^m d_l U_l^m(t), \quad \forall l,
\end{aligned} \tag{5.122}$$

$$\begin{aligned}
\partial\mathcal{L}/\partial I_l(t) &= -d\mu_{3l}(t)/dt = \\
&-\beta_5\left(\frac{2}{C_l^e(t)}\right)\left(\frac{V_l(t)}{C_l^e(t)} - \delta_l\right)\frac{V_l(t)}{I_l(t)} + \sum_{k \in \mathcal{L}} \mu_{1k}(t)\gamma_1 \frac{\partial \hat{V}_k(t)}{\partial I_l(t)} \\
&+ \sum_{k \in \mathcal{L}} \mu_{3k}(t)\gamma_2 \left(\frac{\partial R_k(t)}{\partial I_l(t)}\right) + \mu_{3l}(t)\gamma_2(\xi_1 - \alpha_6 I_l(t))\frac{U_l^c(t)}{C_l^d(t)} \tag{5.123} \\
&+ \mu_4(t)\gamma_1 \frac{\partial \hat{P}(t)}{\partial I_l(t)} + \mu_5(t)\gamma_1 \frac{\partial \hat{K}(t)}{\partial I_l(t)} + \mu_6(t)\gamma_1 \frac{\partial \hat{E}_t}{\partial I_l(t)} \\
&+ \sum_{k \in \mathcal{L}} \omega_{4k}(t) \frac{\partial U_l^m(t)^{\max}}{\partial I_l(t)}, \quad \forall l,
\end{aligned}$$

$$\partial\mathcal{L}/\partial P(t) = -d\mu_4(t)/dt = -\mu_4(t)\gamma_1, \tag{5.124}$$

$$\partial\mathcal{L}/\partial K(t) = -d\mu_5(t)/dt = -\mu_5(t)\gamma_1, \tag{5.125}$$

$$\partial\mathcal{L}/\partial E(t) = -d\mu_6(t)/dt = -\mu_6(t)\gamma_1 - \nu_1(t_1). \tag{5.126}$$

The optimal control is obtained by taking:

$$\frac{\partial\mathcal{L}}{\partial u_p(t)} = 0, \quad \forall p, \tag{5.127}$$

which is:

$$\begin{aligned} \partial \mathcal{L} / \partial U_l^c(t) = & \mu_{2l}(t) - \mu_{3l}(t) \gamma_2 \alpha_6 (I_l(t) - I_p) / C_l^d(t) + \omega_{1l}(t) - \omega_{2l}(t) \\ & + \sum_{k \in \mathcal{L}} \omega_{4k}(t) \frac{\partial U_l^m(t)^{\max}}{\partial U_l^c(t)} - \eta_1(t) c_1^c d_l = 0, \quad \forall l, \end{aligned} \quad (5.128)$$

$$\partial \mathcal{L} / \partial U_l^m(t) = -\mu_{3l}(t) \gamma_2 \frac{\partial M_l(t)}{\partial U_l^m(t)} + \omega_{3l}(t) - \omega_{4l}(t) - \eta_1(t) c_2^m d_l C_l^d(t) = 0, \quad \forall l, \quad (5.129)$$

$$\begin{aligned} \partial \mathcal{L} / \partial U^v(t) = & \sum_{k \in \mathcal{L}} \mu_{1k}(t) \gamma_1 \frac{\partial \hat{V}_k(t)}{\partial U^v(t)} + \sum_{k \in \mathcal{L}} \mu_{3k}(t) \gamma_2 \frac{\partial R_k(t)}{\partial U^v(t)} + \mu_4(t) \gamma_1 \frac{\partial \hat{P}(t)}{\partial U^v(t)} \\ & + \mu_5(t) \gamma_1 \frac{\partial \hat{K}(t)}{\partial U^v(t)} + \mu_6(t) \gamma_1 \frac{\partial \hat{E}(t)}{\partial U^v(t)} + \omega_5(t) - \omega_6(t) = 0, \end{aligned} \quad (5.130)$$

$$\begin{aligned} \partial \mathcal{L} / \partial U_j^p(t) = & \sum_{k \in \mathcal{L}} \mu_{1k}(t) \gamma_1 \frac{\partial \hat{V}_k(t)}{\partial U_j^p(t)} + \sum_{k \in \mathcal{L}} \mu_{3k}(t) \gamma_2 \frac{\partial R_k(t)}{\partial U_j^p(t)} + \mu_4(t) \gamma_1 \frac{\partial \hat{P}(t)}{\partial U_j^p(t)} \\ & + \mu_5(t) \gamma_1 \frac{\partial \hat{K}(t)}{\partial U_j^p(t)} + \mu_6(t) \gamma_1 \frac{\partial \hat{E}(t)}{\partial U_j^p(t)} + \omega_{7j}(t) - \omega_{8j}(t) = 0, \quad \forall j. \end{aligned} \quad (5.131)$$

The Karush-Kuhn-Tucker optimality conditions require that all inequality constraints: $g^j(\vec{x}(t), \vec{u}(t)) \geq 0$, while the Lagrange multipliers: $\omega_j(t) \geq 0$, hence the complementary slackness (CS) conditions:

$$\omega_j(t) g^j(\vec{x}(t), \vec{u}(t)) = 0, \quad \forall j, \quad (5.132)$$

are:

$$\omega_{1l}(t) U_l^c(t) = 0, \quad \forall l, \quad (5.133)$$

$$\omega_{2l}(t) (U_l^{c\max} - U_l^c(t)) = 0, \quad \forall l, \quad (5.134)$$

$$\omega_{3l}(t) U_l^m(t) = 0, \quad \forall l, \quad (5.135)$$

$$\omega_{4l}(t)(U_l^m(t)^{\max} - U_l^m(t)) = 0, \quad \forall l, \quad (5.136)$$

$$\omega_{5l}(t)U^v(t) = 0, \quad \forall l, \quad (5.137)$$

$$\omega_6(t)(U^{v\max} - U^v(t)) = 0, \quad (5.138)$$

$$\omega_{7j}(t)U_j^p(t) = 0, \quad \forall j, \quad (5.139)$$

$$\omega_{8j}(t)(U_j^{p\max} - U_j^p(t)) = 0, \quad \forall j, \quad (5.140)$$

$$\eta_1(t)(F_t^* - \sum_l (c_1^c d_l U_l^c(t) + c_2^m d_l C_l^d(t) U_l^m(t))) = 0. \quad (5.141)$$

Equally for the endpoint emissions, but only at final time $t = t_1$:

$$\nu_1(t_1)(E_{t_1}^* - E(t_1)) = 0. \quad (5.142)$$

The final time to this problem is fixed and there is no final state $S(t_1)$ defined, so the transversality conditions, or boundary points at $t = t_1$, imposed to the costate variables are:

$$\mu_{1l}(t_1) = 0, \quad \forall l, \quad (5.143)$$

$$\mu_{2l}(t_1) = 0, \quad \forall l, \quad (5.144)$$

$$\mu_{3l}(t_1) = 0, \quad \forall l, \quad (5.145)$$

$$\mu_4(t_1) = 0, \quad (5.146)$$

$$\mu_5(t_1) = 0, \quad (5.147)$$

$$\mu_6(t_1) = 0. \quad (5.148)$$

For all the state-variables in $\vec{x}(t)$ the initial states are also known:

$$V_l(t_0) = V_{l0}, \quad \forall l, \quad (5.149)$$

$$C_l^d(t_0) = C_{l0}^d, \quad \forall l, \quad (5.150)$$

$$I_l(t_0) = I_{l0}, \quad \forall l, \quad (5.151)$$

$$P(t_0) = P_0, \quad (5.152)$$

$$K(t_0) = K_0, \quad (5.153)$$

$$E(t_0) = E_0, \quad (5.154)$$

hence a two-point boundary value problem (TBVP), with boundary points specified in equations (5.143) to (5.154), solving simultaneously the first-order state equations (5.114) to (5.119) as well as the first-order costate equations (5.121) to (5.126), while satisfying the zero-order control variables $u_p(t)$ in equations (5.128) to (5.131), as well as (5.133) to (5.142), remains.

The partial derivative expressions in equations (5.121) to (5.131) are given in appendix D.1.

Part III

Strategic modelling examples

Chapter 6

Case studies

6.1 Introduction

In this chapter the modelling framework, models and equations, which have been presented in previous chapters, are applied to some small and medium sized transport networks in order to validate the optimal control model. The results of these case studies will provide more insight into the dynamic model's properties. Furthermore, some potential problems when using Pontryagin's Maximum Principle to solve the models are indicated and discussed.

The different dynamic features to the models are demonstrated using a two link (one route), a three link (three routes) as well as a twelve link (thirty-eight routes) transport network. Before examining the different cases, some further insight in solving the optimal control models using numerical techniques is given.

6.2 Solving the optimal control in Matlab

When ordinary differential equations are required to satisfy boundary conditions at more than one value of the independent variable, the resulting problem is a two-point boundary value problem as seen in paragraph 5.4.2. The two points that need to be satisfied are usually (but not necessarily) the starting and ending values of the integration.

The crucial distinction of two-point boundary value problems with the initial value problems (as for example used to solve problem (A3) in paragraph 5.3.1) is that the boundary conditions at the starting point in the former case do not determine a unique solution to start with, and a 'random' choice among the solutions, which satisfy these incomplete starting boundary conditions, is almost certain not to satisfy the boundary conditions at the other specified point. Hence, some iteration (actually very many) from the initial 'guess' is required to find the unique global solution of the differential equations.

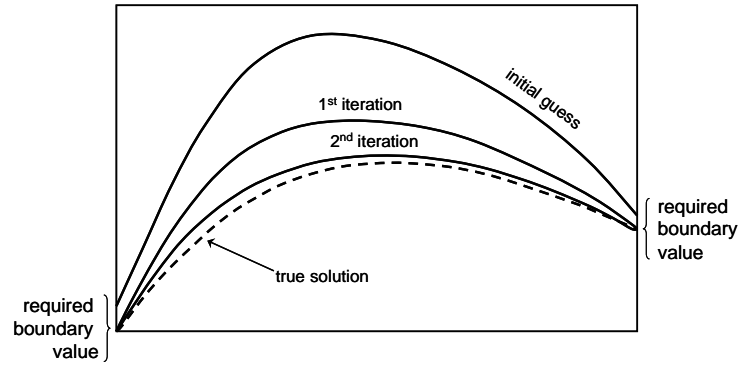


Figure 6.1: Relaxation method (Press et al., 1989).

Generally, two distinct classes of numerical methods for solving two-point boundary value problems exist (Press et al., 1989). In the *shooting method* values for all dependent variables at one boundary are chosen. The ordinary differential equations are then integrated by initial value methods, arriving at the other boundary. The discrepancies with the desired boundary values, for example expressed in a Root Mean Squared Error (RMSE) between N specified and calculated boundary points at time $t = t_b$:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{q=1}^N (y_q^{\text{spec}}(t = t_b) - y_q^{\text{calc}}(t = t_b))^2}, \quad (6.1)$$

are then minimised iteratively using common multidimensional root-finding methods. Hence, this method perfectly resembles the general structure of the open-loop controller depicted before in figure 5.11.

In addition, *relaxation methods* use a different approach, in that the differential equations are replaced by finite difference equations on a mesh of points that cover the range of the integration. A trial solution, derived from the initial guess of dependent variables, then exists at each mesh point, not satisfying the desired finite difference equations, nor necessarily satisfying the required boundary conditions. The relaxation exists of adjusting all the values on the mesh so as to bring them into successively closer agreement with the difference equations, and, simultaneously, with the boundary conditions, as depicted in figure 6.1. Error control in the case of relaxation methods is not done using the RMSE value, but by using the residual $r(t)$ of the continuous solution, which is discussed next.

Such a relaxation method, used here to determine the optimal control with (mixed) (in)equality constraints, is a collocation method named *BVP4c*, and is implemented in MATLAB's problem solving environment (PSE) by Jacek Kierzenka and Lawrence Shampine as reported in Kierzenka and Shampine (2001)

ibid. Shampine et al. (2000). The collocation method directly solves the set of necessary optimality conditions derived from the Maximum Principle applied in previous chapter. This so-called indirect method of solving an optimal control, is known to be efficient, with good convergence properties. A known drawback of this type of methods, however, is that the iterations must start close to a (local) solution in order to solve the problem (Schwartz, 1996).

The function *BVP4c* solves two-point boundary value problems for (systems of) ordinary differential equations, such as those used in this research, i.e.:

$$\frac{dx(t)}{dt} = f(x(t), t), \quad (6.2)$$

with $x(t)$ given at some boundary. It integrates a system of first-order ordinary differential equations on an interval, subject to general two-point boundary conditions. Hence, it produces a solution $x(t)$ that is continuous on the chosen interval and has a continuous first derivative there. *BVP4c* is a finite difference code that implements the 3-stage Lobatto IIIa formula. This is a collocation formula and the collocation polynomial provides a smooth (C1-continuous) solution. Mesh selection and error control are based on the residual of the continuous solution. The collocation is performed with a piecewise cubic polynomial $S(t)$, which is an implicit Runge-Kutta formula with an interpolant (which reduces to the well-known *Simpson method* when applied to a quadrature problem). The continuous solution is fourth-order accurate uniformly in the interval of integration, i.e.: $\|x(t) - S(t)\| = \mathcal{O}(h^4)$, where h is the maximum step size. The error estimation and mesh selection are based on the residual $r(t)$ of $S(t)$, which is:

$$r(t) = \frac{dS(t)}{dt} - f(S(t), t). \quad (6.3)$$

Similarly the residual in the boundary conditions is: $g(S(a), S(b))$. Put differently, $S(t)$ is a solution of the boundary value problem:

$$\frac{dx(t)}{dt} = f(x(t), t) + r(t), \quad (6.4)$$

and:

$$g(y(a), y(b)) = g(S(a), S(b)). \quad (6.5)$$

When the residuals are small, $S(t)$ is supposed to be a good solution.

Collocation techniques use a mesh of points to divide the interval of integration into subintervals. A solver determines a numerical solution by solving a global system of algebraic equations resulting from the boundary conditions, and the

collocation conditions imposed on all the subintervals. Hence, the cubic polynomials collocate at the ends of the subinterval (as well as at the midpoint). The solver then estimates the error of the numerical solution on each subinterval. If the solution does not satisfy the (user-defined) tolerance criteria: $r(t) \leq \varepsilon$, the solver adapts the mesh and repeats the process. An initial mesh as well as an initial approximation of the solution at the mesh points must be provided beforehand.

The function *BVP4c* is used in this research in conjunction with MATLAB¹ to solve the two-point boundary value problem that remains from solving the optimal control model applying the Maximum Principle as derived in previous chapter. Several procedures² are built to obtain an input vector for the problem that is accordingly solved in *BVP4c*. In addition, the control variables, as calculated in the zero-order part of the optimal control model, are derived in separate procedures that are implemented in MATLAB as well, and fed back to the *BVP4c* function. In particular, the modelling of the Karush-Kuhn-Tucker (KKT) multipliers, which are adjoined to the inequality constraints: $g^j(t) \geq 0$:

$$\omega_i(t) = \begin{cases} > 0 & \text{if the constraint is just binding, that is if: } g^j(t) = 0; \\ 0 & \text{if the constraint is non-binding, that is if: } g^j(t) > 0, \end{cases} \quad (6.6)$$

as well as the interaction between several control variables (possibly causing violation of the rank condition, as specified in equation (5.90)), for example through the budget constraint specified in equation (5.76), pose some challenges. Nonlinear programming techniques will have to be applied to solve these zero-order equations. Furthermore, a multi-start feature is built to obtain a better initial guess, which is of great importance when applying this kind of methods to highly nonlinear systems as the ones modelled in this research.

The analytical derivations, including the partial derivatives in appendix D, are derived analytically using the symbolic tool MAPLE³, before being coded and implemented in MATLAB.

6.3 Networks and cases

Three different networks are constructed that will serve as case studies in order to show the different dynamic features, as specified before in chapter 5. The networks are depicted in figures 6.2 to 6.4. The network characteristics have been kept simple, but contain all essential characteristics in order to show the applicability of the optimal control models. The first network consists of one origin-destination pair, connected by two links only ($\mathcal{L} = \{1, 2\}$, $\mathcal{I} = \{1\}$, $\mathcal{J} = \{2\}$, $\mathcal{R}_{12} = \{1\}$). This network is used to demonstrate the general mechanism of control variable interaction. The second network has one origin and two

¹MATLAB version 6.5 release 13.

²The MATLAB m-files with all the coding can be requested from the author.

³MAPLE version 9.

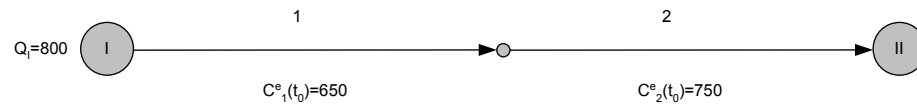


Figure 6.2: Corridor model (network 1).

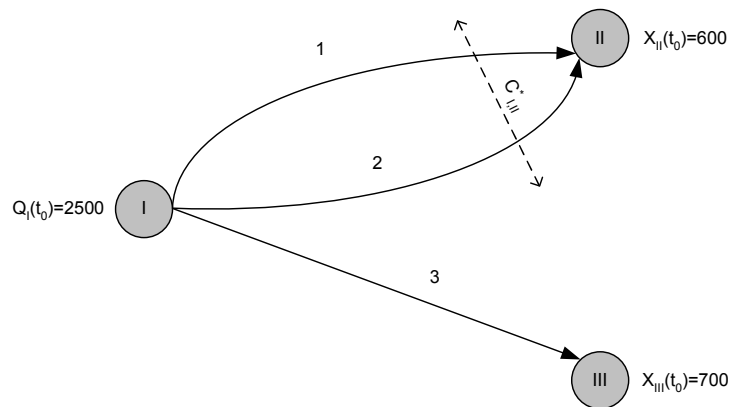


Figure 6.3: Small network model (network 2).

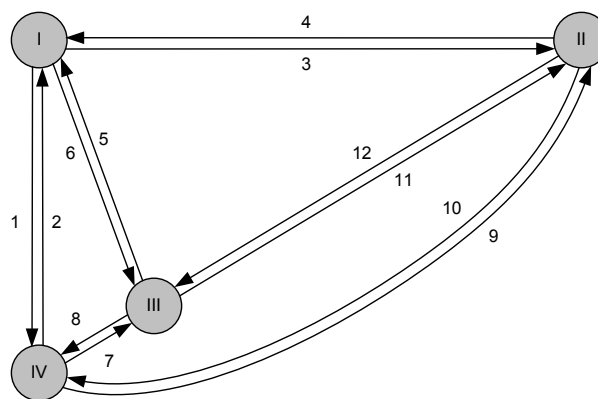


Figure 6.4: Medium sized network model (network 3).

destinations, linked by three routes ($\mathcal{L} = \{1, 2, 3\}$, $\mathcal{I} = \{1\}$, $\mathcal{J} = \{2, 3\}$, $\mathcal{R}_{12} = \{1, 2\}$, $\mathcal{R}_{13} = \{3\}$). This network is initially used to demonstrate the elastic demand as function of the composite costs (shown for $c_{1,2}^*$ in figure 6.3), but also for most of the other cases when applying the different cost criterions. The third network has four origins and four destinations, connected by twelve links ($\mathcal{L} = \{1, 2, \dots, 12\}$, $\mathcal{I} = \{1, 2, 3, 4\}$, $\mathcal{J} = \{1, 2, 3, 4\}$, \mathcal{R}_{ij} in table 6.6), and is used to demonstrate the actual choice behaviour of trip makers in a (strategic) network, particularly for the accessibility cost-criterion.

For each of the case-networks origin-destination - route - link incidence tables, which state the presence or absence of a relationship between mutual network elements and/or social-economic variables, are built. In these tables (shown for the third network only in paragraph 6.4.6), the routes that may comprise an origin-destination pair are specified, not necessary all routes possible, as well as the road links that make up such a route. The different procedures that are modelled in MATLAB use these tables to derive the paths for state $x_i(t)$ and costate variables $\mu_i(t)$ directly (by automatically calculating the vectors of all partial derivatives from paragraph 5.4.3 that are needed for the collocation procedure).

Apart from the network type, the different cases vary in the type of features that are modelled. First of all, they vary in terms of the cost criterion considered, that is: congestion minimisation, accessibility maximisation, person throughput maximisation or equity maximisation. Furthermore, the dynamic features themselves vary in terms of demand elasticity, pavement deterioration mechanism, applied control(s), type of constraints as well as the inclusion of smoothing and penalty terms (the smoothing and penalty terms are discussed in paragraphs 6.4.2 and 6.4.5 respectively). The variation in cases is depicted in table 6.1, while the cases themselves are discussed in subsequent paragraphs.

The different parameters and their values used in the case studies are depicted in table 6.2 (after next page). Most of these values can be found in Donaghy and Schintler (1998), Paterson (1987) and De la Barra (1989), others are guessed or estimated manually⁴.

6.4 Case studies: congestion minimisation (C1)

The first cost-criterion, which is considered in this chapter, is the level-of-service cost criterion, or congestion minimisation cost criterion. This cost-criterion is applied to all three network configurations as well as to a wide variety of cases from an optimal control model with non-elastic travel demand for the smallest network to one with an integral emission constraint, also allowing for elastic

⁴The real estimation of the parameters in the nonlinear dynamic system is not done in the course of this research. Hence, as realistic as possible parameters are used from other research. The lagged-adjustment parameters γ , for example, should at one point be estimated applying statistical signal processing techniques as introduced in appendix B.3.

Table 6.1: Cases

aspect	type	cost criterion			
		congestion C1	access C2	through- put C3	equity C4
demand	A3 ^a non-elastic	1.1 ^b , 2.1			
	A4 elastic	2.2, 2.3, 2.4, 3.1	3.2, 3.3	2.5	2.6
supply	B1 no det.	1.1, 2.1, 2.4, 3.1	3.2, 3.3		
	B2 natural det.	2.2		2.5	2.6
	B3 physical det.	2.3			
controls	capacity	1.1, 2.1, 2.2, 2.3, 2.4, 3.1	3.2, 3.3	2.5	2.6
	maintenance	2.3			
constraints	control	1.1, 2.1, 2.2, 2.3, 3.1	3.2, 3.3	2.5	2.6
	emissions	2.4		2.5	2.6
	int. emissions		3.3		
computing	smoothing	2.1, 2.2, 2.3, 2.4, 3.1	3.2, 3.3	2.5	2.6
	penalty	2.4	3.3		

^ae.g. A3 indicates: travel demand model A3, as discussed in chapter 5.

^be.g. 1.1 indicates: case network 1, model type 1.

travel demand.

The type of case studies, which are considered in this paragraph adopt some of the basic structure formulated in Donaghy and Schintler (1994) *ibid.* Donaghy and Schintler (1998). At points, the simulations in this paragraph also intend to replicate some of the results presented in these papers. However, current investigations revealed this is not a trivial case. Hence, some corrections have been made to the model equations as well as their derivatives. In addition, smoothing terms have been added and simplifications to the models have been introduced as well.

6.4.1 An optimal U^c -control for case 1.1 (B1/C1)

For the first case study, the corridor network, depicted in figure 6.2, is considered. Two links constitute a single route from the origin zone to the destination zone. The travel demand in the origin zone is constant, whereas the objective is to bring the level-of-service on each link to an 80% level, which represents high-density stable traffic flow. One population segment as well as one type of mode are considered. Furthermore, pavement is assumed not to deteriorate due to traffic wear. In addition, only supply-side capacity improvements are the controls.

Summarising, the following system with infrastructure supply model (B1) applied to cost criterion (C1), furthermore assuming: $\delta_l = 0.8$, is considered, which is:

Table 6.2: Parameters and values.

description	symbol	value	dimensions
parameter travel time function	α_1	0.15	[-]
time-value of money parameter	α_2	0.40	[h€ ⁻¹]
conversion parameter IRI to $C_i^e(t)$	α_3	1.74	[km m ⁻¹]
conversion peak hour to an equivalent day	α_4	8.0	[h]
conversion parameter $U_i^m(t)$ to wear $M_i(t)$	α_5	0.2	[cm ⁻¹]
adjustment coefficient	α_6	0.01	[T ⁻¹]
parameter smoothing function J_{smooth}	α_7	var	[-]
parameter travel time function	β_1	4.0	[-]
weight travel time in utility function:			
mode $m=1$ (car)	β_{31}	1.0 ^a	[-]
mode $m=2$ (bus)	β_{32}	var	[-]
weight travel cost in utility function:			
mode $m=1$ (car)	β_{41}	0.4	[h€ ⁻¹]
mode $m=2$ (bus)	β_{42}	var	[h€ ⁻¹]
pavement condition elasticity to $C_i^e(t)$	β_5	0.80	[-]
pavement overlay effectivity	β_7	1.0	[-]
velocity-elasticity of emission for mode m^b :			
mode $m=1$ (car)	$\beta_{81 p}$	0.8	[-]
mode $m=2$ (bus)	$\beta_{82 p}$	0.8	[-]
parameter in cost-criterion (C2)	β_9	0.20	[-]
adjustment coefficient	γ_1	0.25	[T ⁻¹]
adjustment coefficient	γ_2	1.0	[T ⁻¹]
level-of-service (LOS)	δ_i	var	[-]
mode-specific emission factor ^b :			
mode $m=1$ (car)	$\epsilon_{1 p}$	1.65×10^{-2}	[g km ⁻¹]
mode $m=2$ (bus)	$\epsilon_{2 p}$	3.30×10^{-2}	[g km ⁻¹]
reciprocal vehicle-occupancy factor:			
mode $m=1$ (car)	θ_{11}	1.0	[pcu ⁻¹]
mode $m=2$ (bus)	θ_{12}	0.1	[pcu ⁻¹]
vehicle-tax conversion:			
mode $m=1$ (car)	θ_{21}	1.0	[-]
mode $m=2$ (bus)	θ_{22}	1.0	[-]
scale parameter logit model	λ_1	0.3	[h ⁻¹]
scale factor composite cost to induced demand	λ_2	2.0	[h ⁻¹]
scale factor equity cost criterion:			
segmentation $k=1$	$\lambda_{3 1}$	2.0	[h ⁻¹]
segmentation $k=2$	$\lambda_{3 2}$	2.0	[h ⁻¹]
pavement deterioration rate	ξ_1	var	[-]
equivalent standard axel load factor (ESAL):			
average mode	Φ_1	1.4	[-]
parameter penalty function J_{pen} :	χ_1	var	[-]
captive travel demand factor	a_1	0.7	[-]
induced travel demand factor	b_1	0.3	[-]
pavement impact factor	Θ	2.86×10^{-5}	[hpcu ⁻¹]

^aParameter β_{31} is 1.0, but only in case of unimodality and one disutility component.^bOnly one type of pollutant p is considered

$$\begin{aligned}
& \min \int_{t_0}^{t_1} \sum_l \left(\frac{Q_1}{C_l^e(t)} - \delta_l \right)^2 dt \\
& \text{s.t. } \frac{dC_l^e(t)}{dt} = U_l^c(t), \quad \forall l, \\
& \text{and } 0 \leq U_l^c(t) \leq U_l^{c\max}, \quad \forall l, \\
& \text{boundary conditions at } t_0 \text{ and } t_1.
\end{aligned} \tag{6.7}$$

From the system it can be noticed that the travel demand $V_l(t)$ doesn't appear as an equation of motion, because a constant travel demand is assumed, which is: $V_l(t) = Q_1$, while the initial capacity for link 1, $C_1^e(t_0)$, is given a smaller value than that of link 2, $C_2^e(t_0)$. If the maximum control value $U_l^{c\max}$, which is the maximum amount of capacity enhancement per unit of time, is taken equally for both links, the transition paths for the states, costates and multipliers are revealed, as depicted in figures 6.5, 6.7, and 6.8, while the resulting control path and volume-over-capacity path are depicted in figure 6.6.

On the basis of the theory discussed before, one would expect a control path for each link that, given the constant travel demand Q_1 , changes the link capacity level for each link to: $1.25 \times Q_1$, which might take longer to transpire if the initial capacity is lower and/or if the upper control bound is lower. Furthermore, given the linearity of the capacity control, a bang-bang control path, following equation (5.100), may be expected.

Indeed, this bang-bang nature of the optimal control can be observed from these figures, although at the same time it can be seen that the cost-criterion is not minimised at the expected 80% level-of-service, that is volume-over-capacity levels don't fully converge to: $\delta_l = 0.8$. The reason for this is mainly computational. The numerical solver has problems with the heavy switching and finite number of discontinuities in the control path for $U_l^c(t)$. The simulation then also ends with the maximum residual⁵: $r(t) = 3,698.61$, while the requested accuracy is: $r(t) \leq 0.001$.

Though, it can still be seen that given the (non-optimal) control path, the volume-over-capacity levels decrease till levels below 100%, by applying the maximum level of capacity enhancement of $25 \text{pcuh}^{-1} \text{T}^{-1}$ for eight units of time on link 2 and ten units of time on link 1.

In the next case study a smoothing term is introduced to the cost criterion to get around this problem.

⁵Only when the maximum residual doesn't meet the requested accuracy of: $r(t) \leq 0.001$, it will be explicitly mentioned.

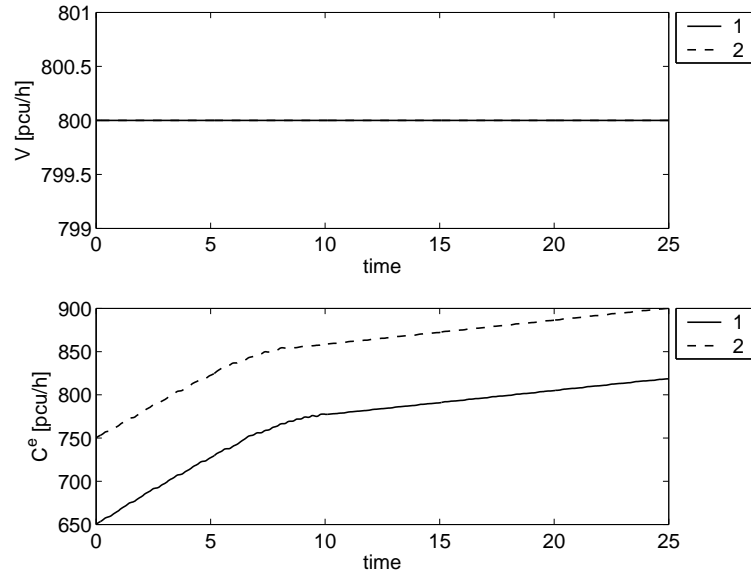


Figure 6.5: States $V_i(t)$ and $C_i^e(t)$ - case 1.1.

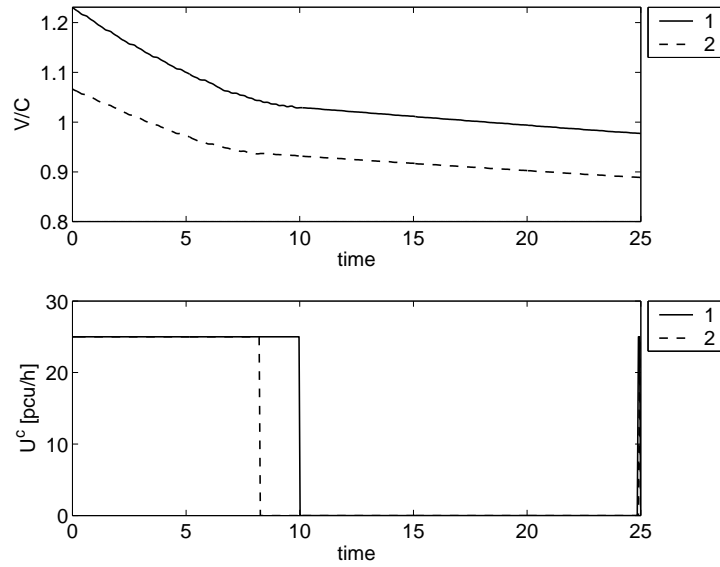


Figure 6.6: Level-of-service $V_i(t)/C_i^e(t)$ and control $U_i^c(t)$ - case 1.1.

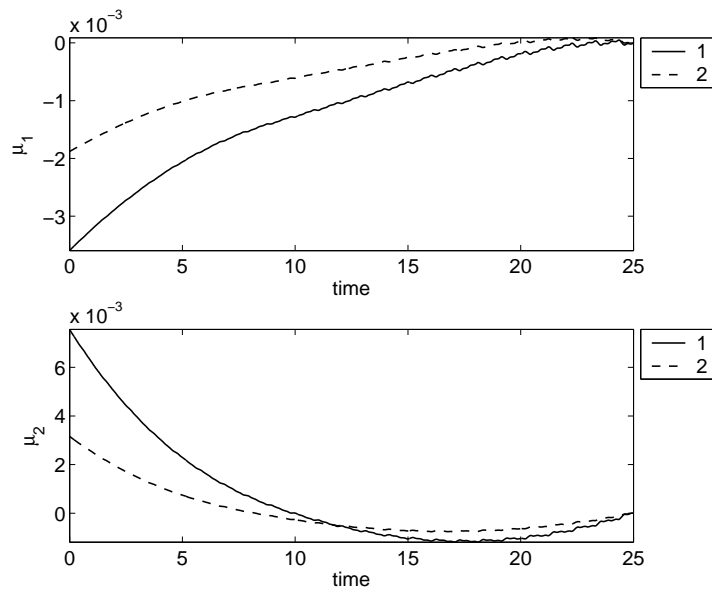


Figure 6.7: Costates $\mu_{1l}(t)$ and $\mu_{2l}(t)$ - case 1.1.

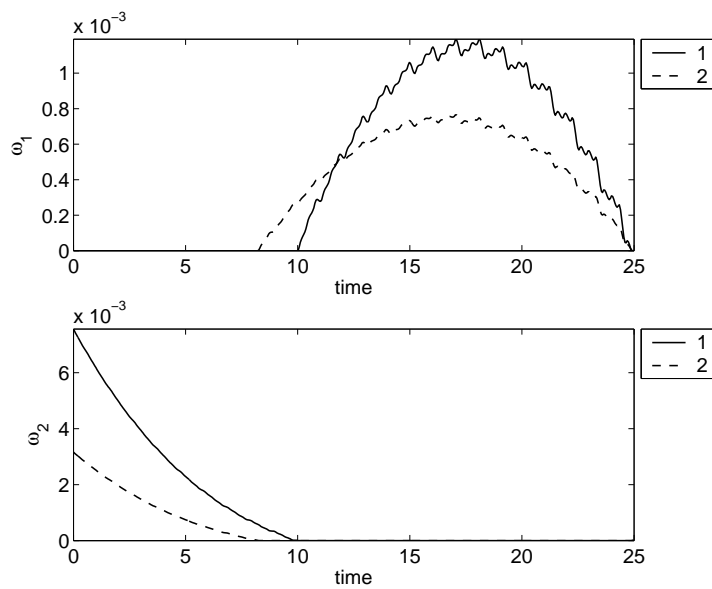


Figure 6.8: Multipliers $\omega_{1l}(t)$ and $\omega_{2l}(t)$ - case 1.1.

Table 6.3: Description of link characteristics (small network).

Link	Length [km]	Free-flow speed [km h ⁻¹]	Capacity at t_0 [pcuh ⁻¹]
1	4	50	650
2	4	50	650
3	4	50	800

6.4.2 An optimal U^c -control for case 2.1 (A3/B1/C1)

For the second case study the small network, depicted in figure 6.3, is considered. Three links make up three routes from one origin zone to two destination zones. The travel demand in the origin zone is, again, constant, whereas the objective is to bring the level-of-service on each link to about 80%, representing high-density stable traffic flow. Different with previous case study is that there is a focus on the route choice mechanism, represented by the disequilibrium travel demand model. Again, one population segment as well as one type of mode are considered. Likewise, pavement is assumed not to deteriorate due to traffic wear, and only supply-side improvements are the controls.

Summarising, the following system with infrastructure supply model (B1) applied to cost criterion (C1), is modelled for the small network, furthermore assuming: $\delta_l = 0.8$, for all three links in the network, which is:

$$\begin{aligned}
& \min \int_{t_0}^{t_1} \sum_l \left(\frac{V_l(t)}{C_l^e(t)} - \delta_l \right)^2 dt \\
& \text{s.t. } \frac{dV_l(t)}{dt} = \gamma_1 \left(\hat{V}_l(t) - V_l(t) \right), \quad \forall l, \\
& \quad \frac{dC_l^e(t)}{dt} = U_l^c(t), \quad \forall l, \\
& \text{and } 0 \leq U_l^c(t) \leq U_l^{c\max}, \quad \forall l, \\
& \quad \text{boundary conditions at } t_0 \text{ and } t_1.
\end{aligned} \tag{6.8}$$

It can be observed from the system that the disequilibrium demand $\hat{V}_l(t)$, which is non-elastic as in equation (5.22), appears as an equation of motion. The utility function $u_{ijmr}(t)$ from equation (5.26), used here, is expressed as a linear equation of route travel time $\tau_r(t)$ only. Note also that in figure 6.3, the attraction values are displayed as non-balanced (remember equation (1.4)). This balancing is done automatically in the coding. Furthermore, the exogenous variables are assumed to be constant over time. A table with a description of the link characteristics is provided as table 6.3.

The computational problem due to linearity of the control variable in the Lagrangean, as observed in previous case study, is addressed here as well, and will be discussed first.

Adding a smoothing term

From the theory in chapter 5, linearity of the Lagrangean and the subsequent bang-bang nature of the control is not a problem as the continuity requirement on the control allows for piecewise continuous functions of time, implying the control is continuous, except possibly at a finite number of points in time. Any discontinuity involves a finite jump, while the state variables $x_i(t)$, costate variables $\mu_i(t)$ as well as the Hamiltonian \mathcal{H} must be continuous, regardless of the discontinuity. Computationally, though, this can be a problem, as the bang-bang solutions cause problems as the Jacobian matrix may become singular on a large domain (see the discussion about this in paragraph 6.4.4). Hence, it is common use to introduce a quadratic smoothing term or perturbed energy term:

$$J_{\text{smooth}}(U_l^c(t)) = \alpha_7 U_l^c(t)^2, \quad \forall l, \quad (6.9)$$

which is adjoined to the cost criterion J , in order to smoothen the discontinuity at the switch-points, as discussed in for example Dadebo et al. (1998) as well as Bertrand and Epenoy (2002).

The general cost-criterion now becomes:

$$J \equiv J + J_{\text{smooth}}(U^c(t)), \quad (6.10)$$

which is:

$$\min \int_{t_0}^{t_1} \sum_l \left(\left(\frac{V_l(t)}{C_l^e(t)} - \delta_l \right)^2 + \alpha_7 U_l^c(t)^2 \right) dt, \quad (6.11)$$

with α_7 given a small positive value. The advantage of adding this term is that $U_l^c(t)$ appears nonlinear in the derivative: $\partial \mathcal{H} / \partial U_l^c(t)$, because the maximised Hamiltonian now reads:

$$\begin{aligned} \mathcal{H} = & - \sum_l \left(\frac{V_l(t)}{C_l^e(t)} - \delta_l \right)^2 - \sum_l \alpha_7 U_l^c(t)^2 + \sum_l \mu_{1l}(t) \gamma_1 \left(\hat{V}_l(t) - V_l(t) \right) + \\ & \sum_l \mu_{2l}(t) U_l^c(t), \end{aligned} \quad (6.12)$$

hence:

$$\frac{\partial \mathcal{H}}{\partial U_l^c(t)} = -2\alpha_7 U_l^c(t) + \mu_{2l}(t), \quad \forall l, \quad (6.13)$$

which is zero if:

$$U_l^c(t) = \frac{\mu_{2l}(t)}{2\alpha_7}, \quad \forall l. \quad (6.14)$$

This implies that the optimal control $U_l^c(t)^*$ can be derived by solving:

$$\frac{\partial \mathcal{L}}{\partial U_l^c(t)} = -2\alpha_7 U_l^c(t) + \mu_{2l}(t) + \omega_{1l}(t) - \omega_{2l}(t) = 0, \quad \forall l, \quad (6.15)$$

constrained by:

$$\omega_{1l}(t) \geq 0, \quad \omega_{2l}(t) \geq 0, \quad \forall l, \quad (6.16)$$

and:

$$(U_l^{c\max} - U_l^c(t)) \geq 0, \quad U_l^c(t) \geq 0, \quad \forall l, \quad (6.17)$$

which is⁶:

$$U_l^c(t)^* = \begin{cases} U_l^{c\max} & \text{and } \omega_{1l}(t) = 0, \quad \omega_{2l}(t) = \frac{\partial \mathcal{H}}{\partial U_l^c(t)}, & \text{if } \frac{\mu_{2l}(t)}{2\alpha_7} \geq U_l^{c\max}; \\ \frac{\mu_{2l}(t)}{2\alpha_7} & \text{and } \omega_{1l}(t) = 0, \quad \omega_{2l}(t) = 0, & \text{if } \frac{\mu_{2l}(t)}{2\alpha_7} < U_l^{c\max}; \\ 0 & \text{and } \omega_{1l}(t) = -\frac{\partial \mathcal{H}}{\partial U_l^c(t)}, \quad \omega_{2l}(t) = 0, & \text{if } \frac{\mu_{2l}(t)}{2\alpha_7} \leq 0. \end{cases} \quad (6.18)$$

Clearly, as: $\partial \mathcal{H} / \partial U_l^c(t) \approx 0$, $U_l^c(t)^*$ gets an intermediate value between the upper and lower bound, hence the function is smoothed around the switch points. By applying a *multi-start* feature that gradually lowers the value of α_7 to almost 0, following Bertrand and Epenoy (2002), the bang-bang structure can be regained in a stepwise manner as good as it gets until the tolerance criterium: $r(t) \leq \varepsilon$, gets violated again. The smoothing term can be seen (in real-life) as a goal for minimising the cost criterion (C1) at the lowest amount of control value $U_l^c(t)$, while aiming at minimised congestion as well.

When applying the smoothing term, on the basis of the theory discussed before, one would expect a control path for each link that, given the link specific traffic demand as derived from the simultaneous distribution, mode choice, route choice model, changes the capacity to the volume-over-capacity level of 80%, taking account of the changing travel behaviour in terms of destination and route-choice of the trip makers in the system. Given the discussion on

⁶Please note that, according to equation (5.100), $U_l^c(t)^*$ is singular when: $\partial \mathcal{H} / \partial U_l^c(t) = 0$. Without giving proof of justification, the lower bound control value has been assigned to this singular component of the control.

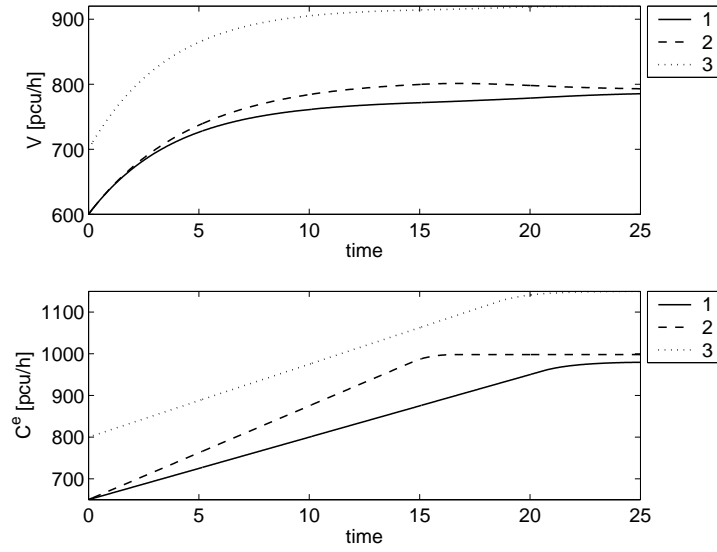


Figure 6.9: States $V_i(t)$ and $C_i^e(t)$ - case 2.1.

the smoothing terms, it is expected that the control path can be characterised partially as having a bang-bang nature, except near the switching points, where the switching is being smoothed.

The results are depicted in figures 6.9 to 6.12. Indeed the transition paths for the state variables, link travel demand $V_i(t)$ and link capacity $C_i^e(t)$, represented in figure 6.9, show the expected behaviour. The initial values (boundary points) for the state variables can be observed from the vertical axes at time: $t = t_0$, and indicate, for the travel demand $V_i(t)$, the deviation from the moving disequilibrium $\hat{V}_i(t)$. The arbitrary chosen initial travel demand is relaxing to an apparent steady-state solution, meanwhile allowing for route-changes in the network to occur. The link capacity is increasing over time as well, due to the continuous addition of new capacity, which can be observed in figure 6.10, representing the values for control values $U_i^c(t)$, with arbitrary chosen (and different) upper bounds $U_i^{c\max}$. The expected bang-bang nature of the optimal control can also be seen, smoothing a bit, mainly because of the value of parameter α_7 , which in this particular case is: $\alpha_7 = 1.0 \times 10^{-6}$. Hence, the volume-over-capacity levels perfectly go down to minimise the cost criterion (C1) at a level of: $\delta = 0.8$, after some units of time.

The costate variables with their transversality boundary conditions: $\mu_{1l}(t_1) = 0$, and: $\mu_{2l}(t_1) = 0$, are depicted in figure 6.11 and clearly show that the required boundary points are met (the RMSE value in equation (6.1) is *de facto* zero when using a collocation method). Most interesting are the values of these costates at the starting point. The open-loop optimal control is completely

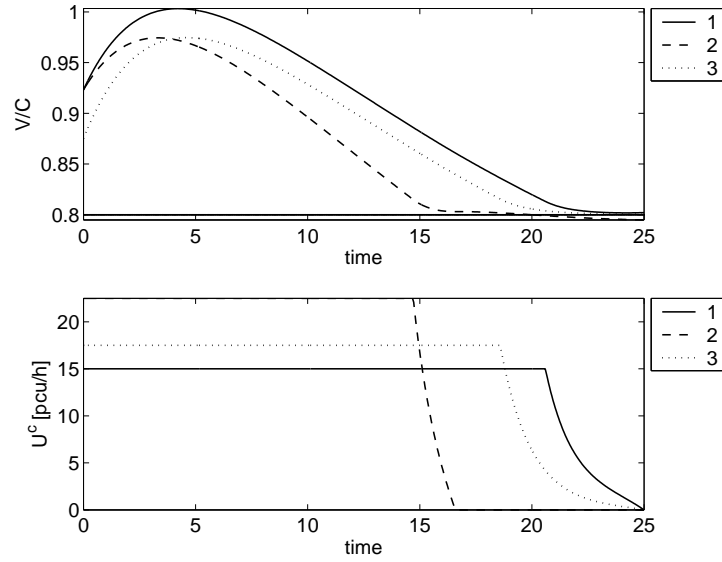


Figure 6.10: Level-of-service $V_i(t)/C_i^e(t)$ and control $U_i^e(t)$ - case 2.1.

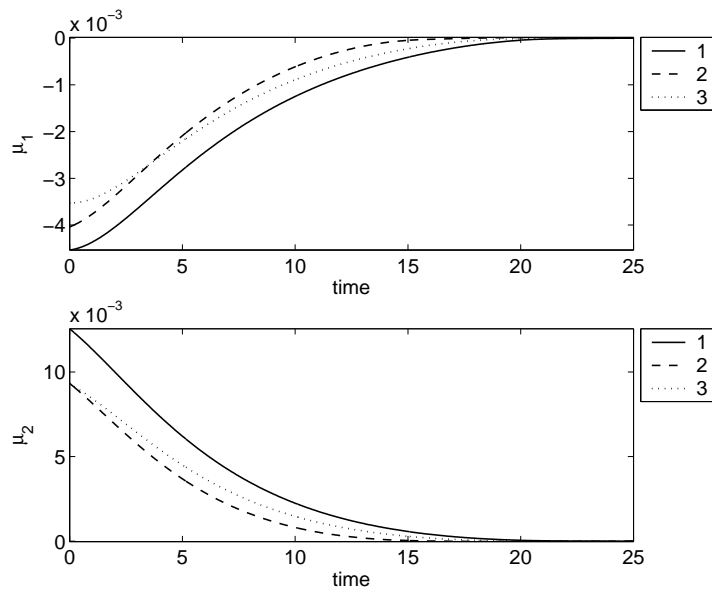


Figure 6.11: Costates $\mu_{1i}(t)$ and $\mu_{2i}(t)$ - case 2.1.

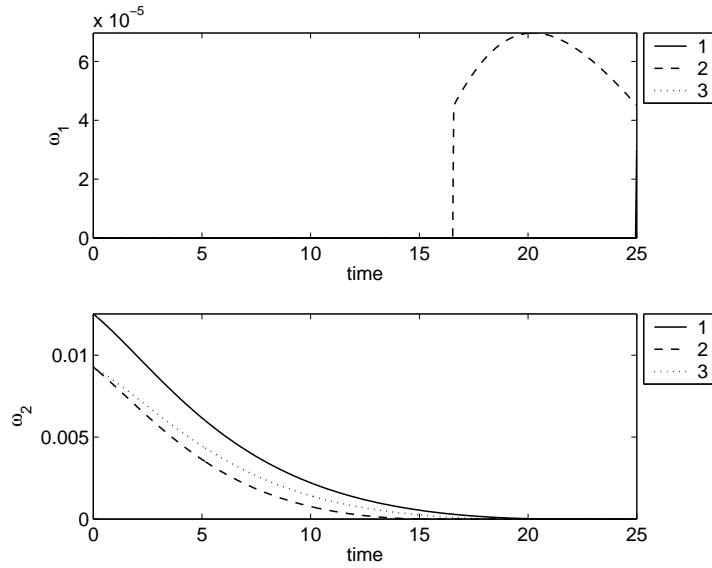


Figure 6.12: Multipliers $\omega_{1l}(t)$ and $\omega_{2l}(t)$ - case 2.1.

determined by these unknowns, which indicate the quality, or economically the ‘shadow price’, whereas the state variables indicate the quantity, since the Lagrangean equation can be seen as having the dimension: quantity \times quality.

Furthermore, the multipliers $\omega_l(t)$ are depicted in figure 6.12 and indicate when upper or lower-bounds for the control $U_l^c(t)$ are met. The values of $\omega_l(t)$ also indicate a shadow price for erecting the boundaries, alike the costate values.

6.4.3 An optimal U^t -control for case 2.2 (A4/B2/C1)

In the third case study the small network is considered again. Whereas in the previous case study the travel demand was non-elastic, it is now considered to be elastic to changes in accessibility (depicted in terms of composite costs on an origin-destination corridor, which is $c_{1,2}^*(t)$ in this particular network). Furthermore, in contrary to previous case study, natural deterioration of the links’ pavement quality is considered. As the objective is, again, to bring the level-of-service on each link to about 80%, representing high-density stable traffic flow, it is expected that the control path for each link changes the link capacity to obtain the envisaged level-of-service level, while taking consideration of induced travel demand as well as pavement deterioration.

Summarising, the following system, combining the dynamic network travel demand model (A4) with infrastructure supply model (B2), while using cost criterion (C1) is modelled. Furthermore, like in the previous case study, a quadratic smoothing term is added, which gives:

$$\begin{aligned}
& \min \int_{t_0}^{t_1} \sum_l \left(\left(\frac{V_l(t)}{C_l^e(t)} - \delta_l \right)^2 + \alpha_7 U_l^t(t)^2 \right) dt \\
& \text{s.t. } \frac{dV_l(t)}{dt} = \gamma_1 \left(\hat{V}_l(t) - V_l(t) \right), \quad \forall l, \\
& \quad \frac{dC_l^e(t)}{dt} = U_l^t(t) - \xi_1 C_l^e(t), \quad \forall l, \\
& \text{and } 0 \leq U_l^t(t) \leq U_l^{t\max}, \quad \forall l, \\
& \quad \text{boundary conditions at } t_0 \text{ and } t_1.
\end{aligned} \tag{6.19}$$

The disequilibrium travel demand $\hat{V}_l(t)$ is elastic as in equation (5.32), while the natural deterioration rate is: $\xi_1 = 0.01$. Again, the utility function $u_{ijmr}(t)$ in equation (5.26) is expressed as a linear equation of route travel time $\tau_r(t)$ only. Realise also that the trip generation variable $Q_i(t)$ for this example, as depicted in figure 6.3, now equals the *potential* trip generation $\hat{Q}_i(t)$, because of the travel demand elasticity.

The simulation results are depicted in figures 6.13 to 6.17. It can be observed (in comparison to previous case study), that, even though the initial state values are the same, the transition paths for the state variables are different. This is caused by the fact that the disequilibrium level of trips is lower, since part of the latent demand is not (yet) induced, hence the state vector paths vary. It is in particular interesting to see that the control paths are different. More capacity enhancement (the construction effort through construction and/or maintenance) is needed to cover the newly induced trips, as well as the deterioration of capacity. Furthermore, it is clear that the control trajectories (bounded by an arbitrary chosen maximum construction effort $U_l^{t\max}$) are of the expected smoothed bang-bang nature, although the smoothing effect for: $\alpha_7 = 1.0 \times 10^{-6}$, is much more visible this time. After some time steps the cost criterion is minimised at: $\delta_l = 0.8$, for all links. The small increase in volume-over-capacity level at the end of the simulation horizon is caused by the fact that there is still pavement deterioration ongoing, whereas the transversality boundary condition: $\mu_{2l}(t_1) = 0$, for all links, brings all control activities to a stop, following equation (6.18).

These costate trajectories are depicted in figure 6.15, and clearly show that the boundary conditions at the endpoint are met. In addition, in figure 6.16, the KKT multipliers for the lower and upper control bounds are shown. The control paths for $U_l^t(t)$ are at the upper-bounds a greater part of the time.

Figure 6.17, in addition, finally shows the total increase of revealed travel demand over time, because of the total increase in system capacity. Also the potential travel demand $Q_i(t)$ is shown, which would be revealed if conditions were improved to free-flow conditions. This is obviously not the case as the cost criterion (C1) aims at a volume-of-capacity level: $\delta_l = 0.8$, for all links. The potential travel demand is assumed to be constant over time.

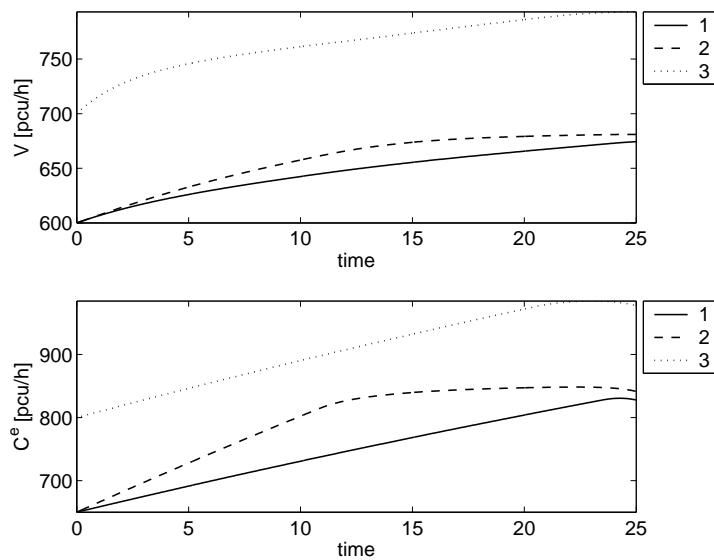


Figure 6.13: States $V_i(t)$ and $C_i^e(t)$ - case 2.2.

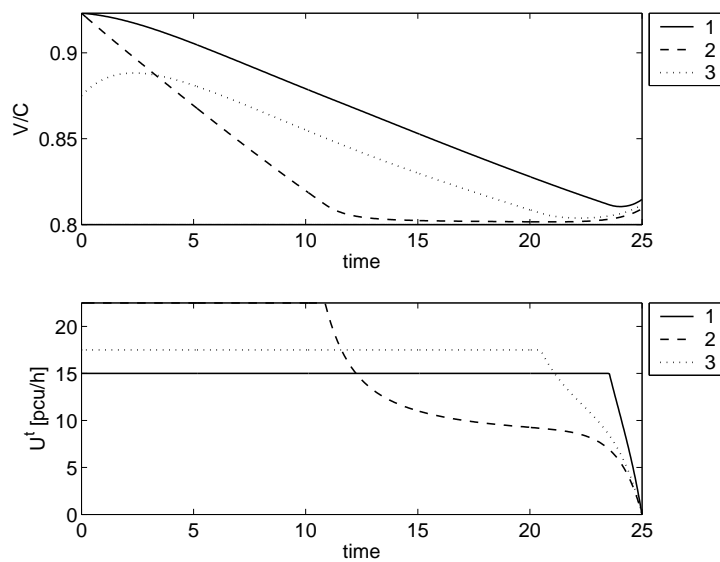


Figure 6.14: Level-of-service $V_i(t)/C_i^e(t)$ and control $U_i^t(t)$ - case 2.2.

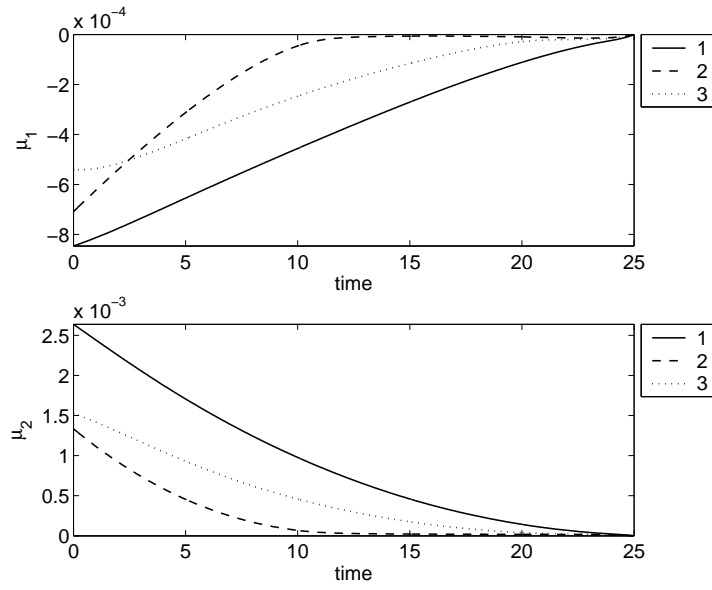


Figure 6.15: Costates $\mu_{1l}(t)$ and $\mu_{2l}(t)$ - case 2.2.

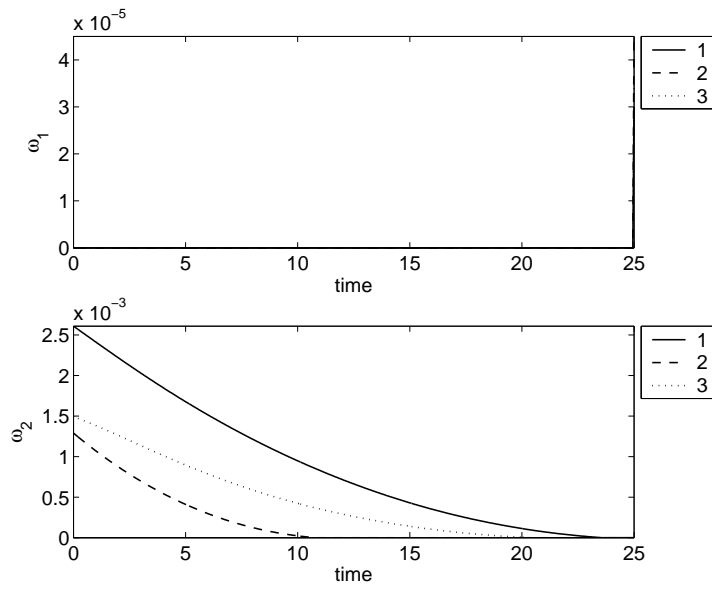


Figure 6.16: Multipliers $\omega_{1l}(t)$ and $\omega_{2l}(t)$ - case 2.2.

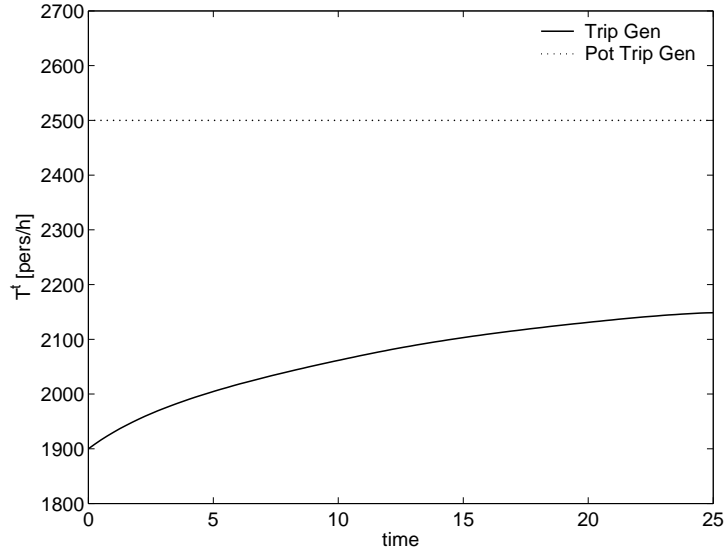


Figure 6.17: (Pot.) trip generation T^t - case 2.2.

6.4.4 An optimal U^c & U^m -control for case 2.3 (A4/B3/C1)

For the fourth case study, the situation with road deterioration due to traffic wear is considered, based on the discussions in chapter 5. Again, the small network is considered, while the travel demand is elastic to changes in network performance. The effective pavement capacity is now dependent on the design capacity (the pavement capacity as initially constructed) as well as the maintenance level expressed through the IRI index. Both the design capacity as well as the maintenance level can be controlled by two separate controls, i.e. design capacity enhancement U^c as well as pavement overlay U^m . The transport policy objective is, again, to bring the level-of-service on each link down to about 80%, hence representing a state of stable high-density traffic flow. It is expected that both controls are active, possibly binding part of the time, in order to bring about the necessary changes, while taking consideration of induced travel demand as well as pavement deterioration.

Summarising, the following system, combining the dynamic network travel demand model (A4) with infrastructure supply model (B3), using cost criterion (C1) is modelled. Furthermore, quadratic smoothing terms are added, this time considering both control variables, i.e.: α_{71} and α_{72} , hence getting:

$$J \equiv J + J_{\text{smooth}}(U^c(t), U^m(t)), \quad (6.20)$$

that is:

$$\begin{aligned}
& \min \int_{t_0}^{t_1} \sum_l \left(\left(\frac{V_l(t)}{C_l^e(t)} - \delta_l \right)^2 + \alpha_{71} U_l^c(t)^2 + \alpha_{72} U_l^m(t)^2 \right) dt \\
& \text{s.t. } \frac{dV_l(t)}{dt} = \gamma_1 \left(\hat{V}_l(t) - V_l(t) \right), \quad \forall l, \\
& \quad \frac{dC_l^d(t)}{dt} = U_l^c(t), \quad \forall l, \\
& \quad \frac{dI_l(t)}{dt} = \gamma_2 \left(R_l(t) + \xi_1 I_l(t) - M_l(t) - \alpha_6 (I_l(t) - I_p) \frac{U_l^c(t)}{C_l^d(t)} \right), \quad \forall l, \\
& \text{and } 0 \leq U_l^c(t) \leq U_l^{c\max}, \quad \forall l, \\
& \quad 0 \leq U_l^m(t) \leq U_l^{m\max}, \quad \forall l, \\
& \quad \text{boundary conditions at } t_0 \text{ and } t_1,
\end{aligned} \tag{6.21}$$

where the disequilibrium demand $\hat{V}_l(t)$ is elastic and the effective capacity is:

$$C_l^e(t) = \alpha_3 \frac{C_l^d(t)}{I_l(t)^{\beta_5}}, \quad \forall l, \tag{6.22}$$

while the maintenance control (equation (5.49)) variable is part of:

$$M_l(t) = \alpha_5 (I_l(t) - I_p) U_l^m(t)^{\beta_7}, \quad \forall l. \tag{6.23}$$

The utility function $u_{ijmr}(t)$ is expressed as a linear equation of route travel time $\tau_r(t)$ only. Furthermore, the upper-level $U_l^m(t)^{\max}$ of equation (5.75), is:

$$U_l^m(t)^{\max} = \left(\frac{I_l(t) - I_p + \gamma_2 \left(R_l(t) + \xi_1 I_l(t) - \alpha_6 (I_l(t) - I_p) \frac{U_l^c(t)}{C_l^d(t)} \right)}{\gamma_2 \alpha_5 (I_l(t) - I_p)} \right)^{\frac{1}{1-\beta_7}}, \quad \forall l. \tag{6.24}$$

Note that $U_l^m(t)^{\max}$ is taken linear here by assuming: $\beta_7 = 1.0$, for simplicity of the calculations. Furthermore, some other simplifications to the model in chapter 5 have been made, that is the dynamic climate factor, road type and modified structural number in equation (5.44) have been replaced by one impact parameter Θ . In addition, there is no natural deterioration, hence the natural deterioration factor is: $\xi_1=0.0$. Reasons for all these simplifications are discussed in the next paragraph. In addition, because this is a one-mode example, an average equivalent axle load factor: $\Phi_1=1.4$, is used, while the smoothing parameters are: $\alpha_{71} = 1.0 \times 10^{-6}$ and: $\alpha_{72} = 2.5 \times 10^{-5}$.

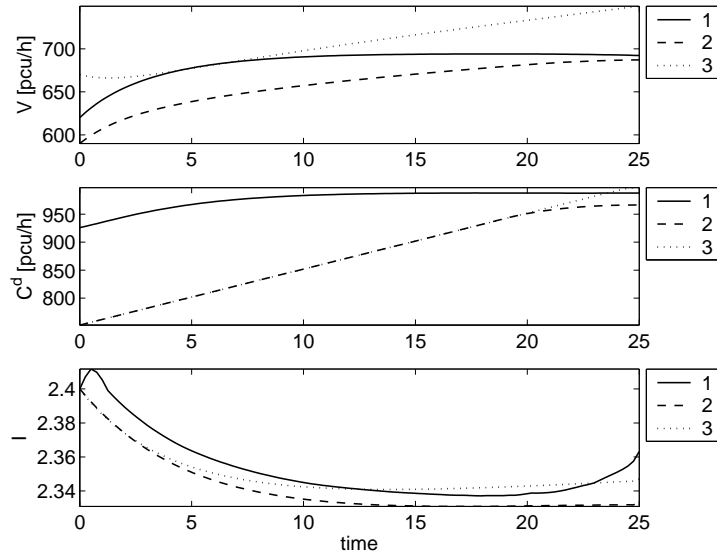


Figure 6.18: States $V_i(t)$, $C_i^d(t)$ and $I_i(t)$ - case 2.3.

Assuming the pavement condition is considered perfect if: $I_i(t) = I_p = 2.0$ m km⁻¹ IRI, while the starting condition for each pavement is equal, but greater than 2.0 m km⁻¹ IRI, the simulation results in figures 6.18 to 6.24 can be obtained. The transition paths of all three state variables show similar patterns as in previous case studies. The pavement condition is improved over time, while the link design capacity as well as the traffic volumes continuously increase. This is also reflected in figure 6.19, where the trajectory for the level-of-service as well as the effective capacity $C_i^e(t)$, conform equation (6.22), is given. Furthermore, in figure 6.20, both controls $U_i^c(t)$ and $U_i^m(t)$ are depicted. For link 3 both controls are at their maximum value most of the time, which is also revealed by looking at the KKT multipliers $\omega_i(t)$ in figures 6.22 and 6.23. Furthermore, it is clear that the maintenance control level is ‘following’ the deterioration due to traffic wear. Total trip generation in this case study is slightly more than in previous case study, as can be seen in figure 6.24. The total number of trips, at the end of the simulation, is about 2140pers⁻¹, which is very close to the 2130pers⁻¹ in the previous case study. This implies that the combination of controls used in this example has about the same effect as the capacity control U^t in previous example alone, while also noting that the value for $U_i^{c\max}$, used here, is lower than $U_i^{t\max}$ in previous case study.

To obtain more realistic results, in particular with respect to the maintenance control, some of the model parameters should be altered. This, however, leads to the problem of *singularities*, which is discussed next.

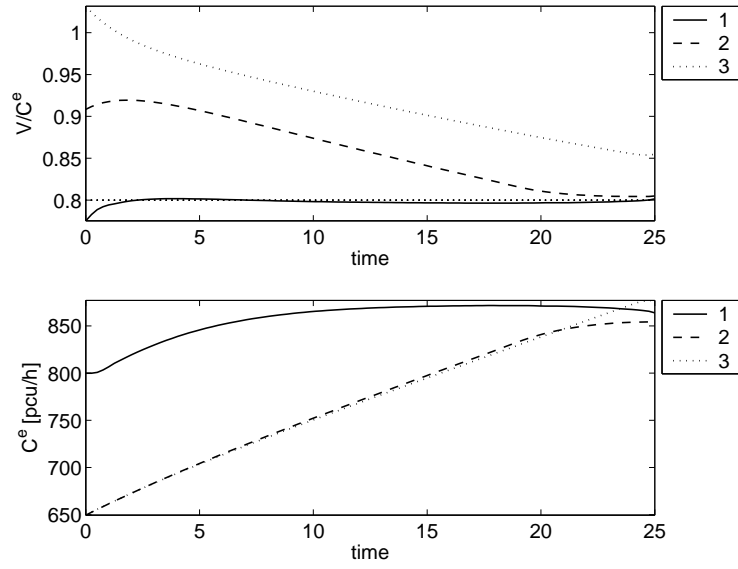


Figure 6.19: Level-of-service $V_i(t)/C_i^e(t)$ and effective capacity $C_i^e(t)$ - case 2.3.

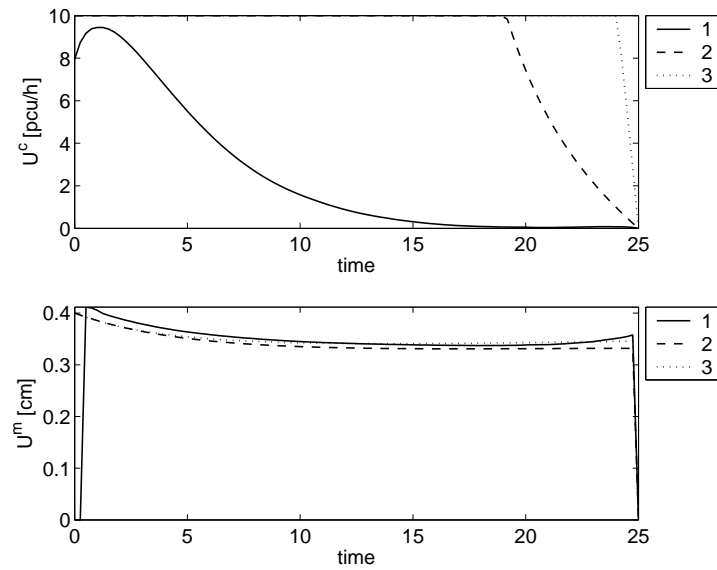


Figure 6.20: Controls $U_i^c(t)$ and $U_i^m(t)$ - case 2.3.

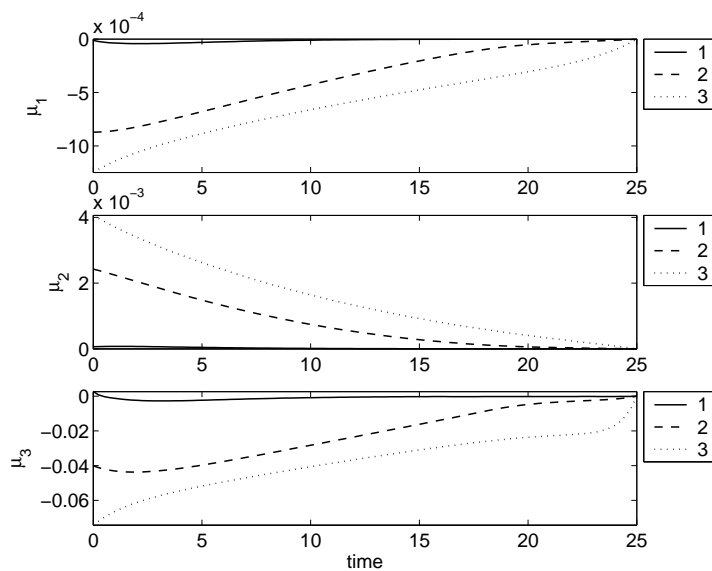


Figure 6.21: Costates $\mu_{1l}(t)$, $\mu_{2l}(t)$ and $\mu_{3l}(t)$ - case 2.3.

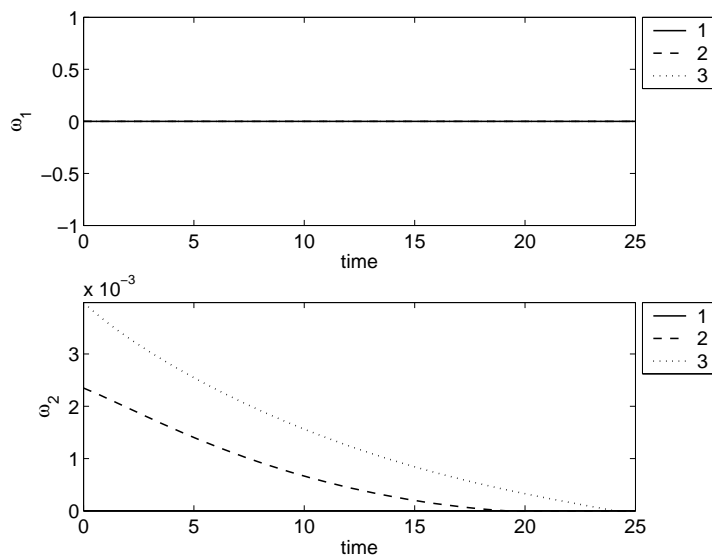


Figure 6.22: Multipliers $\omega_{1l}(t)$ and $\omega_{2l}(t)$ for control $U_l^c(t)$ - case 2.3.

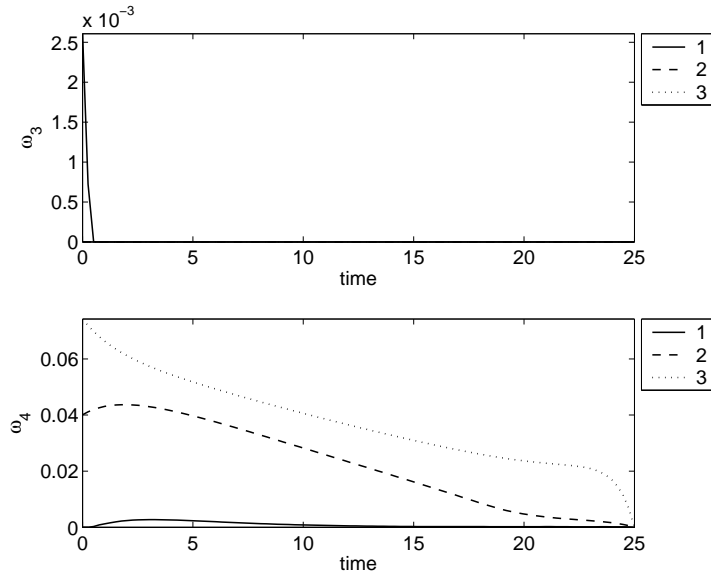


Figure 6.23: Multipliers $\omega_{3l}(t)$ and $\omega_{4l}(t)$ for control $U_l^m(t)$ - case 2.3.

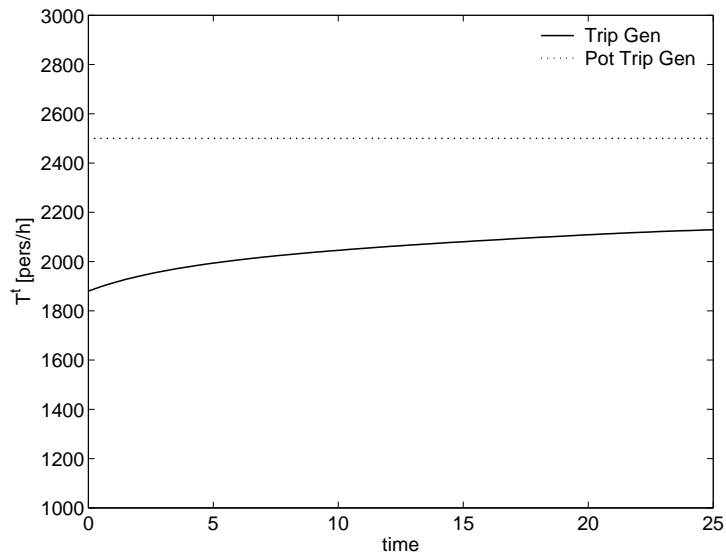


Figure 6.24: (Pot.) trip generation T^t - case 2.3.

Singularity problems

When computing the optimal control, the first problem that arises in the integrating of the system of ordinary differential equations, is a discontinuous function of time, because of the linearity of the control variables in \mathcal{L} . In particular, when the number of bang-bang switches in time is large this is causing computational difficulties. In addition, the convergence theorem's for the method used here, require the equations to be twice continuously differentiable and also require their Jacobian matrix to be non-singular in the vicinity of the solution. Hence, such methods may fail to converge in the case of the non-smooth, highly nonlinear equations as in the models discussed here. This is why quadratic smoothing terms appeared useful.

Stability of nonlinear systems can be checked by obtaining the value of the *Jacobian*, which is the determinant of the *Jacobi matrix*, illustrated here for a system of two nonlinear equations f_1 and f_2 that can be approximated by two Taylor expanded linear functions f_1 and f_2 near (x_1, x_2) , that is:

$$\begin{bmatrix} \frac{\partial f_1(x_1, x_2)}{\partial x_1} & \frac{\partial f_1(x_1, x_2)}{\partial x_2} \\ \frac{\partial f_2(x_1, x_2)}{\partial x_1} & \frac{\partial f_2(x_1, x_2)}{\partial x_2} \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix} = - \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix}, \quad (6.25)$$

which can also be written as:

$$\mathcal{J} \Delta x = -f, \quad (6.26)$$

with Δx an array of the small linear expansions in h and k . The 2×2 matrix is the Jacobi matrix \mathcal{J} .

Generalised for a $n \times n$ matrix, equation (6.26), solved for f becomes:

$$\begin{bmatrix} x_1(m+1) \\ x_2(m+1) \\ \vdots \\ x_n(m+1) \end{bmatrix} = \begin{bmatrix} x_1(m) \\ x_2(m) \\ \vdots \\ x_n(m) \end{bmatrix} - \mathcal{J}^{-1}(x_1, \dots, x_n) \begin{bmatrix} f_1(x_1, \dots, x_n) \\ f_2(x_1, \dots, x_n) \\ \vdots \\ f_n(x_1, \dots, x_n) \end{bmatrix}, \quad (6.27)$$

which is also known as Newton's method for a $n \times n$ system. Because of its size it quickly grows with the problem's size. The Jacobi matrix is not constant, but varies as the numerical iterations in the solver proceed. The Jacobi matrix J is *singular* if it cannot be inverted, that is if the determinant (called the Jacobian):

$$\det \mathcal{J} \equiv |\mathcal{J}| = 0. \quad (6.28)$$

For the models described in previous chapter, it is generally very difficult to analyse the Jacobian, as the models are rather complex (at least in their size).

Leaving out the elastic demand $\hat{V}_l(t)$, hence assuming constant trip generation capacity Q , as in model (A1), the singularity problem is illustrated for the remaining model. When assuming a simple construction and maintenance control (B3) as well as cost criterion (C1), for one link only, that is:

$$\begin{aligned}
& \min \int_{t_0}^{t_1} \left(\frac{V(t)}{C^e(t)} - \delta \right)^2 + \alpha_{71} U^c(t)^2 + \alpha_{72} U^m(t)^2 dt \\
& \text{s.t. } \frac{dV(t)}{dt} = \gamma_1 (Q - V(t)), \\
& \quad \frac{dC^d(t)}{dt} = U^c(t), \\
& \quad \frac{dI(t)}{dt} = -U^m(t), \\
& \text{and } 0 \leq U^c(t) \leq U^{c\max}, \\
& \quad 0 \leq U^m(t) \leq U^{m\max}, \\
& \quad \text{boundary conditions at } t_0 \text{ and } t_1,
\end{aligned} \tag{6.29}$$

with:

$$C^e(t) = \alpha_3 \frac{C^d(t)}{I(t)^{\beta_5}}, \tag{6.30}$$

also noting that both (non-bounded) control equations can be written as, similar to equation (6.18), as:

$$\frac{dC^d(t)}{dt} = \frac{\mu_2(t)}{2\alpha_{71}}, \tag{6.31}$$

and:

$$\frac{dI(t)}{dt} = -\frac{\mu_3(t)}{2\alpha_{72}}. \tag{6.32}$$

The 6×6 Jacobi matrix \mathcal{J} now reads (leaving out time in the notation):

$$\left[\begin{array}{cccccc} -\gamma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2\alpha_{71}} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{2\alpha_{72}} \\ \frac{2(I^{\beta_5})^2}{\alpha_3^2 C^{d^2}} & -\frac{2V(I^{\beta_5})^2}{\alpha_3^2 C^{d^3}} & \frac{2V(I^{\beta_5})^2 \beta_5}{\alpha_3^2 C^{d^2} I} + \gamma_1 & 0 & 0 & 0 \\ -\frac{2V(I^{\beta_5})^2}{\alpha_3^2 C^{d^3}} & \frac{2(\frac{V}{C^e} - \delta) I^{\beta_5}}{\alpha_3 C^{d^2}} + \frac{\alpha_3 C^{d^2}}{2V^2 (I^{\beta_5})^2} & \frac{2(\frac{V}{C^e} - \delta) I^{\beta_5} \beta_5}{\alpha_3 C^{d^2} I} & 0 & 0 & 0 \\ \frac{2(\frac{V}{C^e} - \delta) I^{\beta_5}}{\alpha_3 C^{d^2}} & \frac{4(\frac{V}{C^e} - \delta) V I^{\beta_5}}{\alpha_3 C^{d^3}} & \frac{2(\frac{V}{C^e} - \delta) V I^{\beta_5} \beta_5}{\alpha_3 C^{d^2} I} & 0 & 0 & 0 \\ \frac{2V(I^{\beta_5})^2 \beta_5}{\alpha_3^2 C^{d^2} I} + \frac{2(\frac{V}{C^e} - \delta) I^{\beta_5} \beta_5}{\alpha_3 C^{d^2} I} & -\frac{2V^2 (I^{\beta_5})^2 \beta_5}{\alpha_3^2 C^{d^3} I} - \frac{2(\frac{V}{C^e} - \delta) V I^{\beta_5} \beta_5}{\alpha_3 C^{d^2} I} & \frac{2V^2 (I^{\beta_5})^2 \beta_5^2}{\alpha_3^2 C^{d^2} I^2} + \frac{2(\frac{V}{C^e} - \delta) V I^{\beta_5} \beta_5^2}{\alpha_3 C^{d^2} I^2} & 0 & 0 & 0 \end{array} \right] \quad (6.33)$$

with the Jacobian being:

$$|\mathcal{J}| = \gamma_1^2 V^2 (I^{\beta_5})^2 \beta_5 (2V^2 (I^{\beta_5})^2 \beta_5 - 3V I^{\beta_5} \beta_5 \delta \alpha_3 C^d - 3V^2 (I^{\beta_5})^2 + 5V I^{\beta_5} \delta \alpha_3 C^d + \delta^2 \alpha_3^2 C^{d^2} \beta_5 - 2\delta^2 \alpha_3^2 C^{d^2}) / (\alpha_3^4 C^{d^6} I^2 \alpha_{71} \alpha_{72}), \quad (6.34)$$

which is zero at the roots for the numerator (solved for I), that is if:

$$I = 0 \quad \vee \quad I = \left(\frac{\delta \alpha_3 C^d}{V} \right)^{\frac{1}{\beta_5}} \quad \vee \quad I = \left(\frac{\delta \alpha_3 C^d (-2 + \beta_5)}{V(2\beta_5 - 3)} \right)^{\frac{1}{\beta_5}}. \quad (6.35)$$

In a contour plot, figure 6.25 (next page), these roots may be visualised for different values of travel demand V . The iso-lines in this plot indicate the combinations of design capacity C^d and maintenance level I , given a value for V , where the Jacobian becomes singular, hence these are points where the numerical solver fails (in the case of equation (6.29)). A close look at these lines, indicates that, unfortunately, the points where the Jacobian gets singular are non-trivial, which implies that they are positioned in the ‘common’ solution space.

Related problems when deriving an optimal U^v & U^p -control

In the examples discussed till now, the control variables $U_l^c(t)$ and $U_l^m(t)$ have been linear or linearised. The control variables U^v and U^p , as presented in equation (5.26), where utility is expressed as a function of both route travel time $\tau_r(t)$ as well as route travel cost $\kappa_{ijmr}(t)$, although linear in the utility function, are embedded in the nonlinear demand elasticity model $D_{ij}(t)$ as well as the discrete-choice model $G_{ijmr}(t)$.

In order to solve this problem the zero-order equation (5.97) should be solved. This equation requires that:

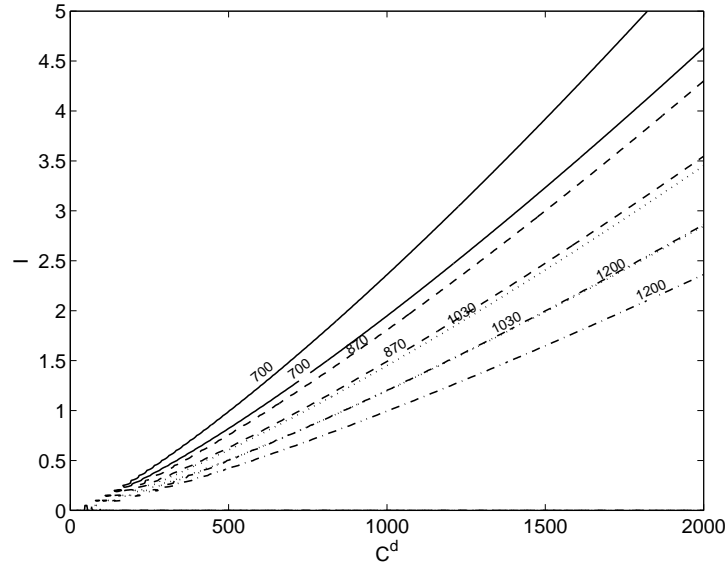


Figure 6.25: Contour plot for singular Jacobian (at different values of V).

$$\frac{\partial \mathcal{L}}{\partial U^v(t)} = \frac{\partial \mathcal{H}}{\partial U^v(t)} + \omega_5(t) - \omega_6(t) = 0, \quad (6.36)$$

and:

$$\frac{\partial \mathcal{L}}{\partial U_j^p(t)} = \frac{\partial \mathcal{H}}{\partial U_j^p(t)} + \omega_{7j}(t) - \omega_{8j}(t) = 0, \quad \forall j, \quad (6.37)$$

in combination with the KKT constraints (6.16) and (6.17). However, the partial derivative terms $\partial \mathcal{H} / \partial U^v(t)$ and $\partial \mathcal{H} / \partial U_j^p(t)$ are complicated functions of the control variables, for which solving is not a trivial case anymore. These partial derivatives are given in equations (5.130) and (5.131) as well as in appendix D.1.

To solve this zero-order problem, which can be written as a system of nonlinear equations in conjunction with the collocation method and some routines in MATLAB, should be possible, but is not done here. In general this can be done by applying root-finding techniques (finding x that satisfies: $f(x) = 0$) or optimisation techniques (finding x that satisfies: $df/dx = 0$). As any optimisation method can be used as a root-finding method by asking it to find the minimum of the absolute value of the function, and since nonlinear equation solving is much harder than nonlinear optimisation, it is probably best to apply some line search method for multidimensional unconstrained optimisation problems.

For example, applying the multivariate Newton-Raphson method (which requires the inverse of the Jacobian, hence there is fear for above mentioned singularity problems), a sequential quadratic programming approach (which is a generalisation of Newton's method) or the unconstrained optimisation Downhill Simplex Method (which doesn't require derivatives), all described in Press et al. (1989)⁷.

The importance of applying an algorithm that somehow finds the root for these controls is stressed in the concluding chapter 7.

6.4.5 An optimal U^c -control with an emission state constraint for case 2.4 (A3/B1/C1)

For the fifth case study, the small network is considered once again. The inputs to this case study are similar to those in the third case study. This implies that an elastic travel demand model is considered, which is sensitive to changes in network performance (accordingly translated into changes in composite costs). Furthermore, the objective remains equal in aiming for an 80% level-of-service, implying high-density stable traffic flow. Even though this case doesn't consider pavement deterioration, the major difference with the third case study is that an emission state constraint, alike equation (5.82), is applied.

By doing so, one would expect the control path to change in comparison with case study three, in order to minimise the cost criterion as much as possible, while taking consideration of the elastic demand, but not violating the emission constraint. Remember that on urban roads, the emission factor is monotonously decreasing with increasing speeds, as in equation (3.25), which can be accomplished by capacity enhancement for example.

To add a pure state inequality constraint to the model is not a trivial case, because such a constraint, by definition, doesn't contain a control variable. Even though in chapter 5 formal methods for incorporating such constraints by directly adjoining them to the Lagrangean are discussed, it is common use (especially for numerical optimal control problems), to use a penalty function method, which is discussed next.

Adding a state constraint using a penalty function

Application of an inequality endpoint state constraint, in the sense of equation (5.107), is difficult to implement if the system is non-autonomous, as is the case with the dynamic transport model. In particular, if the endpoint constraint is also not directly dependent on the value of one of the control variables, as is the case in equation (5.86). Constraints exhibiting this property are called pure state variable inequality constraints. If the constraint is formulated like a continuous state constraint, compare for example equation (5.82), which is

⁷The MATLAB optimisation toolbox and some internet sources provide computational procedures for nonlinear programming.

the case here, a penalty function or barrier function in the objective function or cost criterion can be applied, alike the smoothing term for the control variables. Doing so, the constrained problem, with respect to the state constraint, is transformed into an unconstrained problem.

In general a penalty function J_{pen} that is adjoined to the cost criterion J as to represent mixed inequality constraints: $g^j(x_i(t), u_p(t)) \leq 0$, (see system 5.87) or pure inequality state constraints: $g^j(x_i(t)) \leq 0$, has the property:

$$J_{\text{pen}}(t) = \begin{cases} 0 & \text{if } \max_j g^j(x_i(t)) \leq 0, j = 1, 2, \dots, m'; \\ \gg 0 & \text{if } \max_j g^j(x_i(t)) > 0, j = 1, 2, \dots, m'. \end{cases} \quad (6.38)$$

Two examples of common penalty functions, the first one being applied here, are, following Van den Boom and De Schutter (2004):

$$J_{\text{pen}}(x) = \chi_1 \sum_j \max(0, g^j(x_i(t)))^2, j = 1, 2, \dots, m', \quad (6.39)$$

and:

$$J_{\text{pen}}(x) = \max_j \max(0, e^{\chi_1 g^j(x_i(t))} - 1)^2, j = 1, 2, \dots, m', \quad (6.40)$$

with: $\chi_1 \gg 1$ ⁸. Assuming $g^j(x_i(t))$ is smooth, both penalty functions have the advantage that the derivative is continuous. For more details on incorporating constraint functions into the objective function, the reader should consult Bryson and Ho (1975) or Van den Boom and De Schutter (2004).

Implementing penalty function (6.39) as $J_{\text{pen}}(E(t))$ for the pure state constraint⁹ (5.82), this reads:

$$g(E(t)) = (E(t) - E^*) \leq 0, \quad (6.41)$$

which is illustrated for an arbitrary: $\chi_1 = 3$, alongside with penalty function (6.40) in figure 6.26. Note that if: $(E(t) - E^*) = 0$, the constraint is just-binding. Hence, the penalty function is not useful if the true optimum is on the boundary, since the calculated optimum might be infeasible, because it always lies outside the feasible region. In that case it is better to use barrier functions (see also Van den Boom and De Schutter, 2004), although numerically, barrier functions are (even) less stable than the penalty function.

⁸In some cases it might be necessary to just allow: $\chi_1 > 0$, or to make χ_1 very large, because the penalty term value needs to be of the order of the rest of functional value J , see for example paragraph 6.5.2.

⁹Penalty function (6.40), at first sight, would give better results, but numerical instability proved to be a lesser problem with penalty function (6.39).

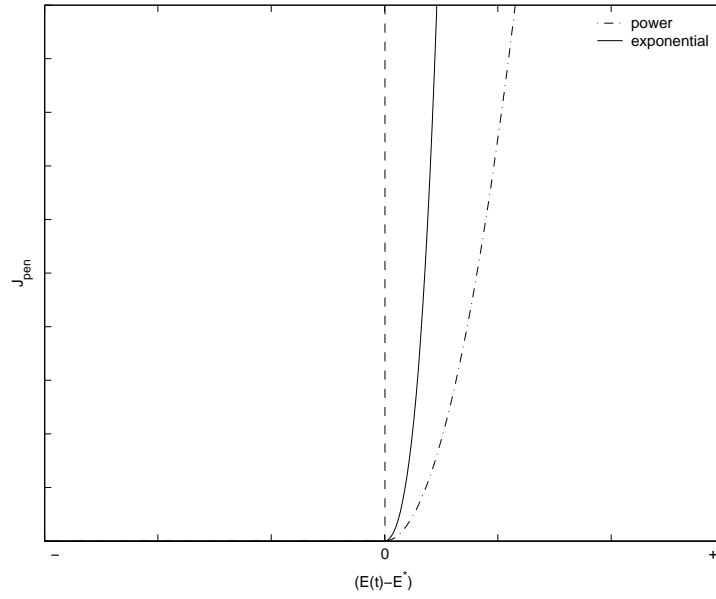


Figure 6.26: Penalty functions for: $(E(t) - E^*) \leq 0$.

Going back to the case study itself, the following system has to be solved:

$$\begin{aligned}
 & \min \int_{t_0}^{t_1} \sum_l \left(\left(\frac{V_l(t)}{C_l^e(t)} - \delta_l \right)^2 + \alpha_7 U_l^c(t)^2 + \chi_1 \max(0, (E(t) - E^*))^2 \right) dt \\
 & \text{s.t. } \frac{dV_l(t)}{dt} = \gamma_1 (\hat{V}_l(t) - V_l(t)), \quad \forall l, \\
 & \quad \frac{dC_l^e(t)}{dt} = U_l^c(t), \quad \forall l, \\
 & \quad \frac{dE(t)}{dt} = \gamma_1 (\hat{E}(t) - E(t)), \\
 & \text{and } 0 \leq U_l^c(t) \leq U_l^{c\max}, \quad \forall l, \\
 & \quad \text{boundary conditions at } t_0 \text{ and } t_1.
 \end{aligned} \tag{6.42}$$

Note that a disequilibrium equation for the total emissions $E(t)$ has been added. The parameters for the disequilibrium emissions $\hat{E}(t)$ are representing the emission factor function for volatile organic compound (VOC) emissions, which is equation (3.26), and serves as a proxy for the total emissions in the system (hence, this could be one of the other pollutants $p \in \mathcal{P}$ as well). Furthermore, the environmental capacity is assumed to be: $E^* = 81$ kg, which is arbitrary chosen within the feasible solution region. As before, the travel demand $\hat{V}_l(t)$ is elastic, whereas the utility is a linear function of route travel time $\tau_r(t)$ only.

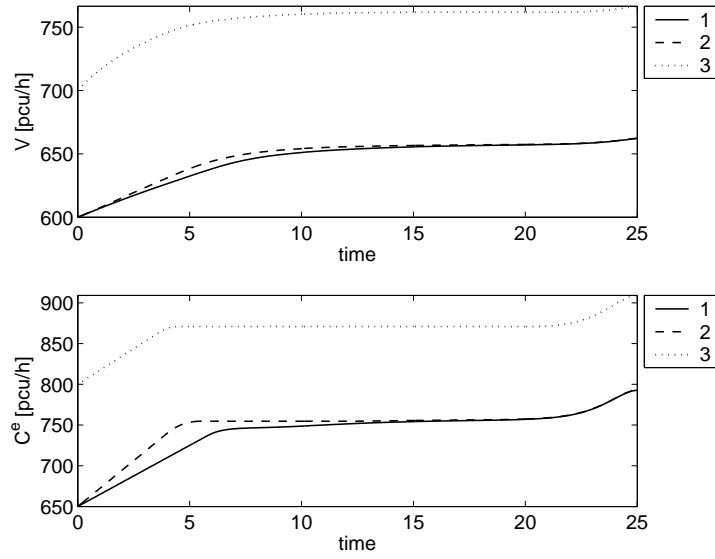


Figure 6.27: States $V_i(t)$ and $C_i^c(t)$ - case 2.4.

Furthermore, note that the cost criterion now comprises of a smoothing term as well as penalty term. In other words:

$$J \equiv J + J_{\text{smooth}}(U^c(t)) + J_{\text{pen}}(E(t)). \quad (6.43)$$

The smoothing parameter is: $\alpha_7 = 1.0 \times 10^{-6}$, whereas: $\chi_1 = 0.05$. The simulation results are depicted in figures 6.27 to 6.33. The transition paths for the state variables follow a similar path in comparison to previous case studies. However, from the volume-over-capacity ratios in figure 6.28, it can be noticed that the cost criterion is not being minimised at: $\delta_l = 0.8$, but slightly above that, which is obviously caused by the pure state constraint, that forces the control path to be as depicted in figure 6.28. The control paths for $U_i^c(t)$ switch between lower and upper bounds, showing the bang-bang nature of the model, although the smoothing effect of α_7 is slightly visible as well.

The costate paths are depicted in figures 6.29 and 6.32, and clearly show that the boundary conditions at the endpoint are met. In addition, in figure 6.30, the KKT multipliers for the lower and upper bounds are shown, through which the bang-bang nature of the control is nicely illustrated.

Figure 6.31 shows the transition paths for the system effects. Alike the general character of the penalty function, the target E^* is slightly exceeded, in particular near the endpoint, where the open-loop control forces $\mu_2(t)$ down to zero, implying zero construction. Exceeding this target, apparently, is only justified for a short period in time.

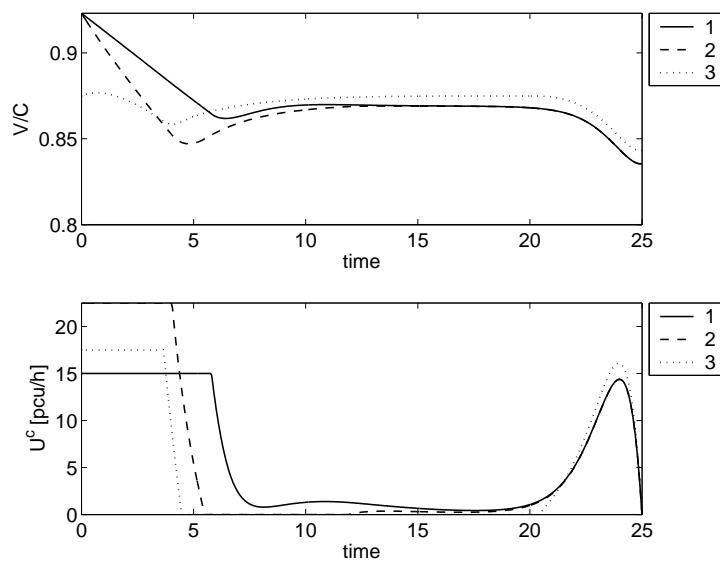


Figure 6.28: Level-of-service $V_i(t)/C_i^e(t)$ and control $U_i^c(t)$ - case 2.4.

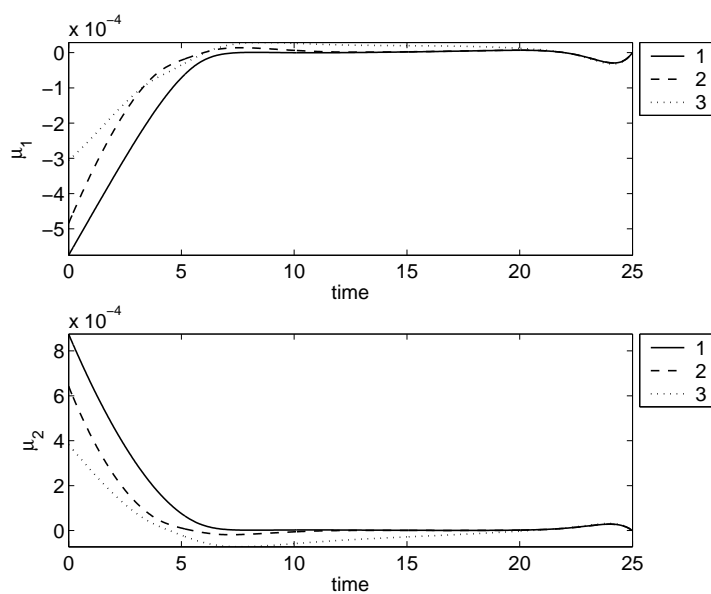


Figure 6.29: Costates $\mu_{1l}(t)$ and $\mu_{2l}(t)$ - case 2.4.

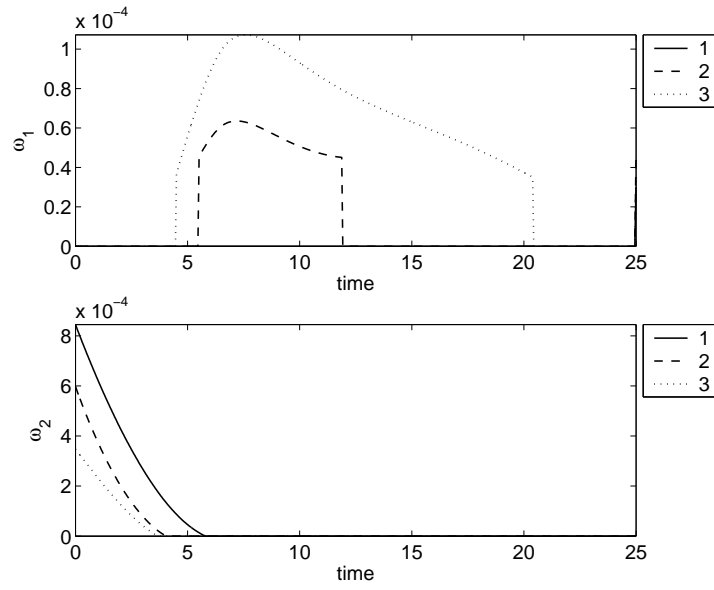


Figure 6.30: Multipliers $\omega_{1l}(t)$ and $\omega_{2l}(t)$ - case 2.4.

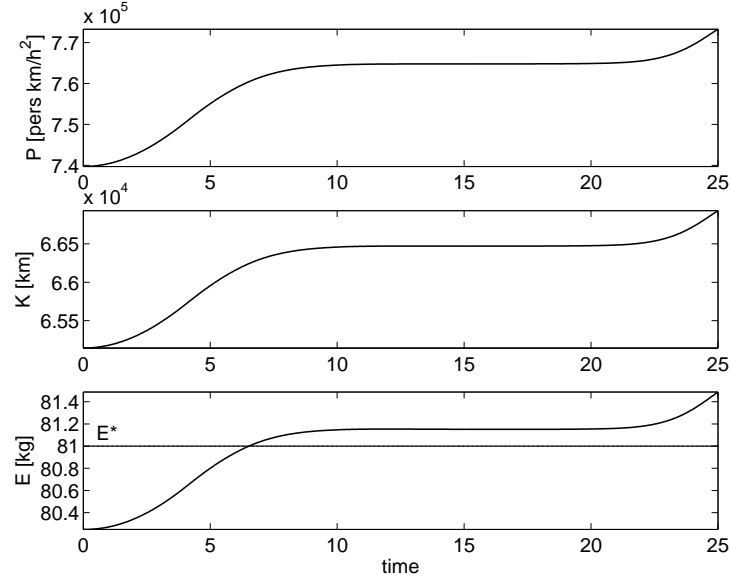


Figure 6.31: States $P(t)$, $K(t)$ and $E(t)$ - case 2.4.

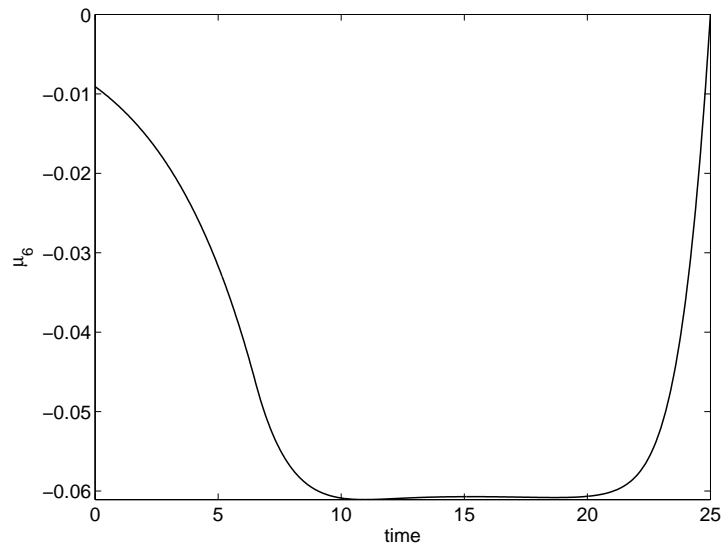


Figure 6.32: Multiplier μ_6 - case 2.4.

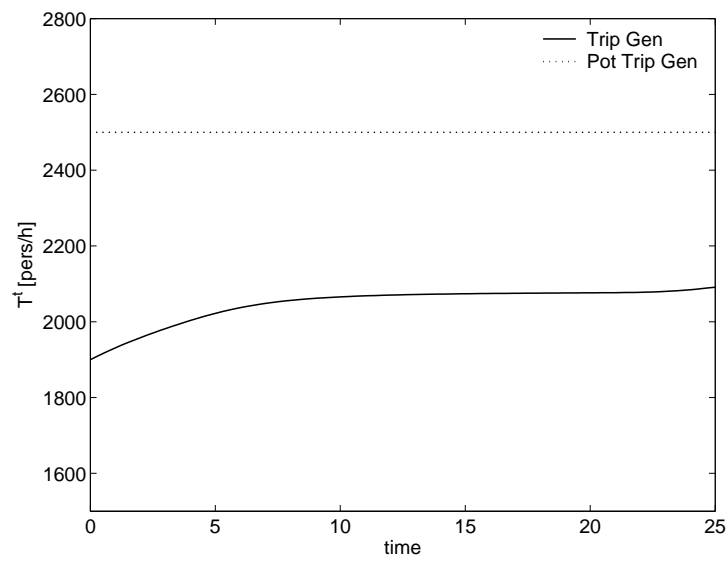


Figure 6.33: (Pot.) trip generation T^t - case 2.4.

Table 6.4: Description of link characteristics (medium network).

Link	Length [km]	Free-flow speed [km h ⁻¹]	Capacity at t ₀ [pcuh ⁻¹]
1	6	50	700
2	6	50	700
3	6	50	800
4	6	50	700
5	5	50	700
6	5	50	800
7	1	50	1100
8	1	50	1100
9	9	50	800
10	9	50	700
11	9	50	700
12	9	50	800

Finally, figure 6.33 shows the total increase of revealed travel demand over time to a level very close to 2100pers h^{-1} , because of the elasticity function $\hat{V}_i(t)$. Also the potential travel demand $\hat{Q}(t)$ is shown, which would be revealed if conditions were improved to free-flow conditions and the environmental capacity was set sufficient high. This is obviously not the matter in this case study. Furthermore, the potential travel demand is assumed to be constant over time. In comparison, it can be shown that without the pure state constraint total pollution would exceed 83kg , whereas total throughput would be over 2160pers h^{-1} . A marginal difference for this example, but it shows the effect of a state constraint. The total pollution level for this example can be brought back to about 79kg at a total throughput level of 1985pers h^{-1} (only captive travellers), which is at: $D_{ij} = 0.7$, for both origin-destination combinations in this case study.

6.4.6 An optimal U^c -control for case 3.1 (A3/B1/C1)

For the sixth case study, the medium sized network, depicted in figure 6.4 is considered. In this network twelve links make up thirty-eight routes from four origins to four destinations. Three tables with respectively a description of link characteristics in table 6.4, social-economic characteristics of the zones in table 6.5, as well as origin-destination-link-route incidence information in table 6.6, from which the route-sets \mathcal{R}_{ij} and \mathcal{R}_{ij}^l can be easily derived, are given. As can be observed from table 6.5, the exogenous social-economic characteristics are assumed to be constant over time in this case study.

The optimal control derived for the medium network is the same as the model (A3/B1/C1) in the third case study, see paragraph 6.4.3. However, this time also equations of motion for the system induced effects, i.e. person throughput $P(t)$ from equation (5.54), distance travelled $K(t)$ from equation (5.57) as well as total emissions $E(t)$ from equation (5.59), are added to the system. Furthermore, the values for the parameters are: $\alpha_7 = 5.0 \times 10^{-6}$ and: $\delta_l = 0.8$.

Table 6.5: Constant social-economic characteristics (medium network).

Zone	Trips [pers h ⁻¹]	t ₀	t ₁	...	t ₂₅
1	$\hat{Q}_1(t)$	2500	2500	...	2500
	$X_1(t)$	500	500	...	500
2	$\hat{Q}_2(t)$	2000	2000	...	2000
	$X_2(t)$	600	600	...	600
3	$\hat{Q}_3(t)$	2500	2500	...	2500
	$X_3(t)$	700	700	...	700
4	$\hat{Q}_4(t)$	2000	2000	...	2000
	$X_4(t)$	600	600	...	600

To monitor the throughput in number of generated person trips [pers h⁻¹], an extra equation of motion is also adjoined to the system, that is:

$$\frac{dT^t(t)}{dt} = \gamma_1 \left(\hat{T}^t(t) - T^t(t) \right), \quad (6.44)$$

with:

$$\hat{T}^t(t) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} (\hat{Q}_i(t) D_{ij}(t) G_{ijmr}(t)). \quad (6.45)$$

These equations are necessary because an initial guess: $T^t(t_0) = T_0^t$, is hard to obtain for this particular network. In general, it can be said that a disadvantage of using systems of differential equations in a boundary-value problem is that in the search process for an optimal solution only the vector with all variable information remains known. This implies that extra information one might want to have, like total number of generated trips (which is all known during the optimisation process), can only be stored through the introduction of new state variables. Likewise, if one wants to obtain a modal split on a certain link, the mode-specific link volumes need to be known. This can only be done by introducing state variables for each mode, as discussed before with the system induced effects.

The initial guess for the traffic volumes $V_l(t_0)$ as well as the initial number of generated person trips, are therefore obtained using a user-equilibrium traffic assignment in a standard transport model¹⁰, assuming an average demand elasticity factor: $D_{ij} = 0.85$.

The simulation results are depicted in figures 6.34 to 6.40. It can be observed that some link capacities are expanded, whereas others are not. Also, the upper-bounds for the control $U_l^c(t)$ are not binding for those expanded links throughout the simulation, hence the value of the derivative $\partial \mathcal{H} / \partial U_l^c(t)$ determines the value of the control for some time. It can be seen that the cost

¹⁰OMNITRANS version 4.0.

Table 6.6: OD-link-route incidence (medium network).

OD	Link	Routes	OD	Link	Routes	OD	Link	Routes
12	01	03, 04	23	01		34	01	29
	02			02			02	
	03	01		03			03	
	04			04	17		04	
	05			05			05	29
	06	02		06	17		06	
	07	04		07	16		07	
	08			08			08	27
	09	03		09			09	
	10			10	16		10	28
	11	02, 04		11			11	28
	12			12	15		12	
13	01	06	24	01	20	41	01	
	02			02			02	30
	03	07		03			03	
	04			04	20		04	32
	05			05			05	31
	06	05		06			06	
	07	06		07			07	31
	08			08	19		08	
	09			09			09	32
	10			10	18		10	
	11			11			11	
	12	07		12	19		12	
14	01	08	31	01		42	01	
	02			02	22		02	35
	03	10		03			03	35
	04			04	23		04	
	05			05	21		05	
	06	09		06			06	
	07			07			07	33
	08	09		08	22		08	
	09			09			09	34
	10	10		10			10	
	11			11	23		11	33
	12			12			12	
21	01		32	01		43	01	
	02	13, 14		02			02	38
	03			03	25		03	
	04	11		04			04	
	05	12		05	25		05	
	06			06			06	38
	07			07			07	36
	08	14		08	26		08	
	09			09	26		09	37
	10	13		10			10	
	11			11	24		11	
	12	12, 14		12			12	37

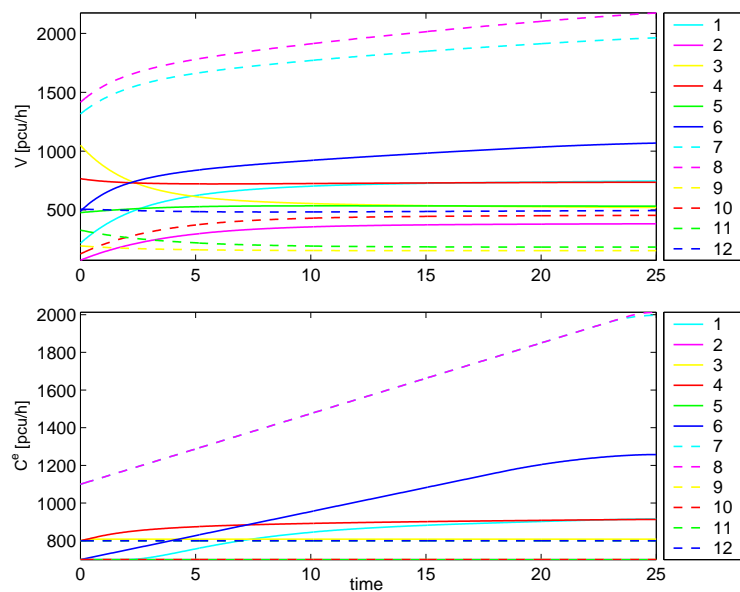


Figure 6.34: States $V_i(t)$ and $C_i^e(t)$ - case 3.1.

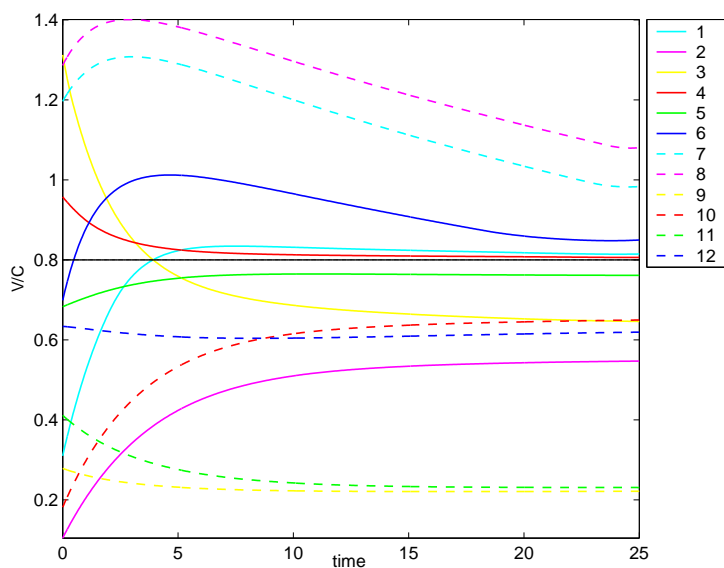


Figure 6.35: Level-of-service $V_i(t)/C_i^e(t)$ - case 3.1.

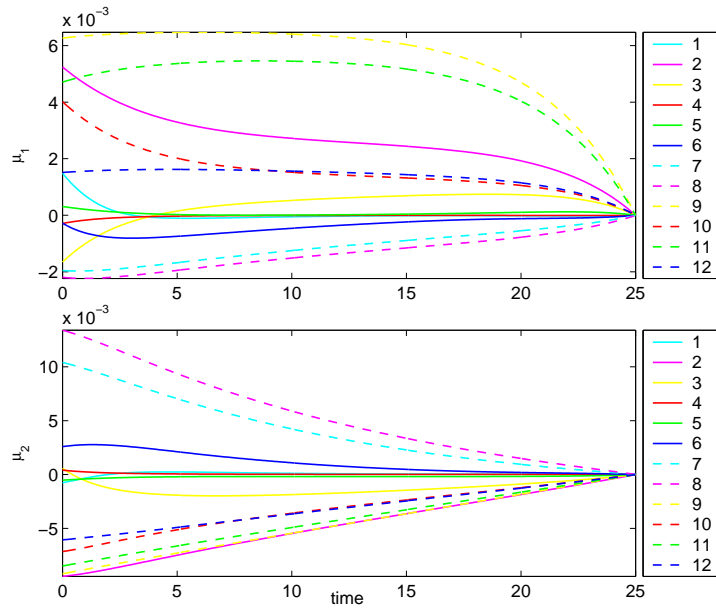


Figure 6.36: Costates $\mu_{1i}(t)$ and $\mu_{2i}(t)$ - case 3.1.

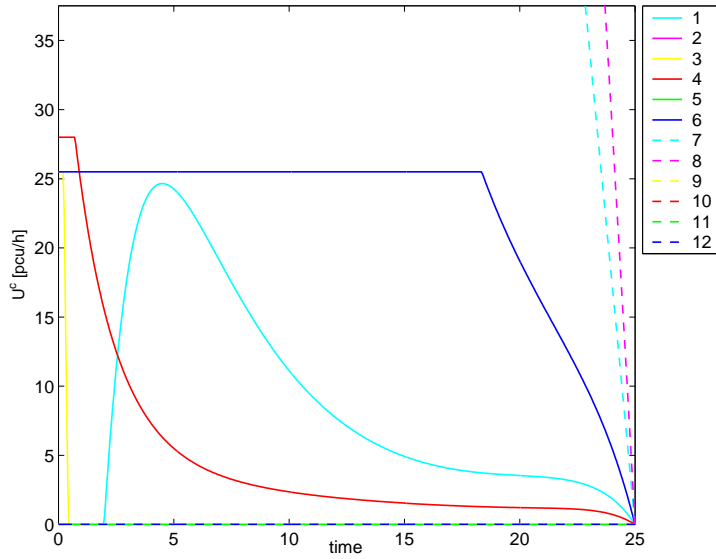


Figure 6.37: Control $U_i^c(t)$ - case 3.1.

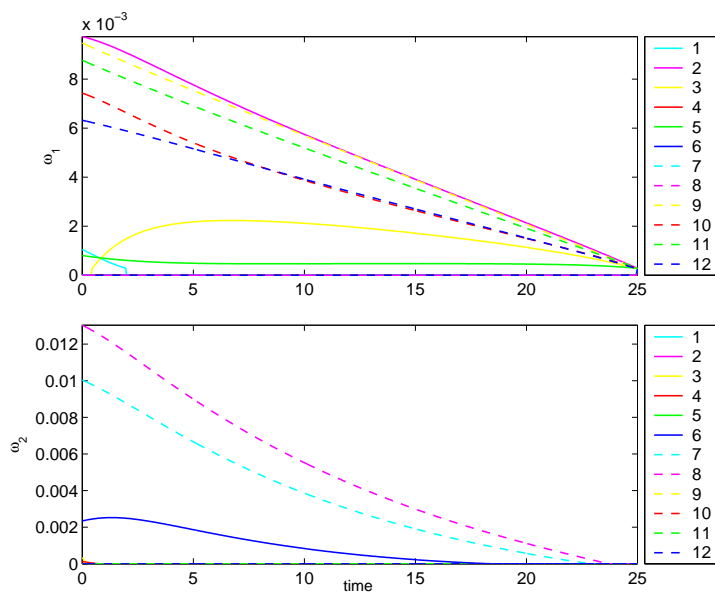


Figure 6.38: Multipliers $\omega_{1l}(t)$ and $\omega_{2l}(t)$ - case 3.1.

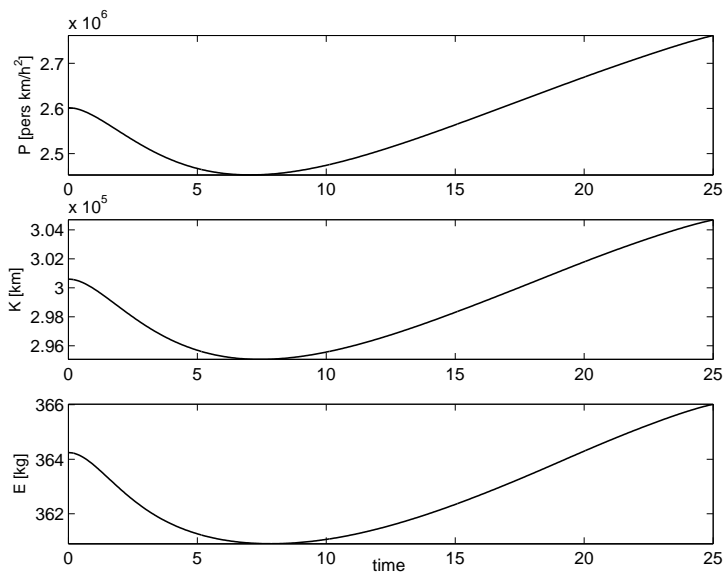


Figure 6.39: States $P(t)$, $K(t)$ and $E(t)$ - case 3.1.

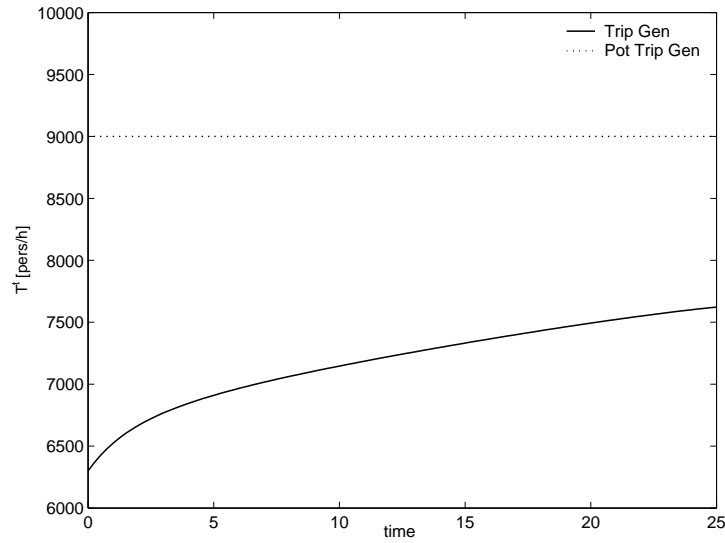


Figure 6.40: (Pot.) trip generation T^t - case 3.1.

criterion is not minimised at volume-over-capacity ratios: $\delta_l = 0.8$. This because negative capacity enhancement is assumed to be impossible, hence some links remain at lower levels of δ_l , whereas for other links maximum construction is applied to lower the value of δ_l in the direction of an 80% level. In addition, the system induced effects show increasing levels of pollution and vehicle-kilometres driven, as well as person throughput levels. The initial decline in these effects is caused by the lagged-adjustment of the state variables to the changed conditions in the network. It is also obvious that the starting conditions, which are derived exogenously, are an overestimation compared to the disequilibrium levels. Hence, the initial guess of: $D_{ij} = 0.85$, could perhaps have been lower. Furthermore, as the induced effects are not controlled, their multipliers (not shown) $\mu_5(t)$, $\mu_6(t)$ and $\mu_7(t)$ are zero throughout the simulation. The overall level of trip making is increasing to a little over 7600 persh⁻¹ at an emission level of 365 kg. Besides, in this case the total integral emissions are: 9430 kg (not shown).

6.5 Case studies: accessibility maximisation (C2)

The second cost criterion, which is considered in this chapter, is the one of accessibility maximisation. Remember that this cost criterion refers to the ease of movement between zones in conjunction to the (relative) attractiveness of these zones, following Sales Filho's definition of accessibility. The cost criterion is applied to the medium network only, as the criterion is best illustrated in a network with several origins and destinations, connected through corridors with different route options (hence different composite cost 'bundles' connecting

centres of activity).

Two case studies applying this cost criterion are discussed next, one without an environmental capacity constraint and one with. Both case studies are based on the dynamic model allowing for travel demand elasticity. Parameter values and the initial guess for state variables are equal to previous case study, where the first cost criterion has been applied to the medium network.

6.5.1 An optimal U^c -control for case 3.2 (A3/B1/C2)

As said before, in this seventh case study the medium sized network is considered again, amongst others using the incidence information depicted in table 6.6, which is similar to previous case study, except for the cost criterion that is considered. Summarising, the dynamic network travel demand model (A3) in combination with infrastructure supply model (B1), using cost criterion (C2) is modelled, which is:

$$\begin{aligned}
 & \max \int_{t_0}^{t_1} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \left(\tilde{Q}_i(t) \tilde{X}_j(t) \right)^{\frac{1}{2}} \exp(-\beta_9 c_{ij}^*(t)) - \sum_l \alpha_{71} U_l^c(t)^2 dt \\
 & \text{s.t. } \frac{dV_l(t)}{dt} = \gamma_1 \left(\hat{V}_l(t) - V_l(t) \right), \quad \forall l, \\
 & \quad \frac{dC_l^e(t)}{dt} = U_l^c(t), \quad \forall l, \\
 & \text{and } 0 \leq U_l^c(t) \leq U_l^{c\max}, \quad \forall l, \\
 & \quad \text{boundary conditions at } t_0 \text{ and } t_1,
 \end{aligned} \tag{6.46}$$

where the disequilibrium demand $\hat{V}_l(t)$ is elastic as in equation (5.22), while the utility is a linear function of route travel time $\tau_r(t)$ only. Again a smoothing term for the control $U_l^c(t)$ is adjoined to the cost criterion, with: $\alpha_7 = 1.0 \times 10^{-6}$.

Besides, four equations of motion are added to the system to monitor the person throughput $P(t)$, kilometres travelled $K(t)$, total emissions $E(t)$ as well as total number of generated person trips $T^t(t)$. Given the formulation of the cost criterion, knowing that the value of α_7 is low, one would expect the control path to be at the maximum control value all the time. This because there are no bounds (like an environmental constraint) to the allocation of these resources, even though for some origin-destination pairs the control is more beneficial (implying their contribution to the cost criterion J is relatively large), which can be read from the shadow prices $\mu_1(t)$ and $\mu_2(t)$. Because of this, more emissions and more trips in the system may be expected, as the accessibility is improved throughout the system and throughout time.

The simulation results are depicted in figures 6.41 to 6.47. Indeed, it can be observed that, for the given parameter values, the control paths are at the value of

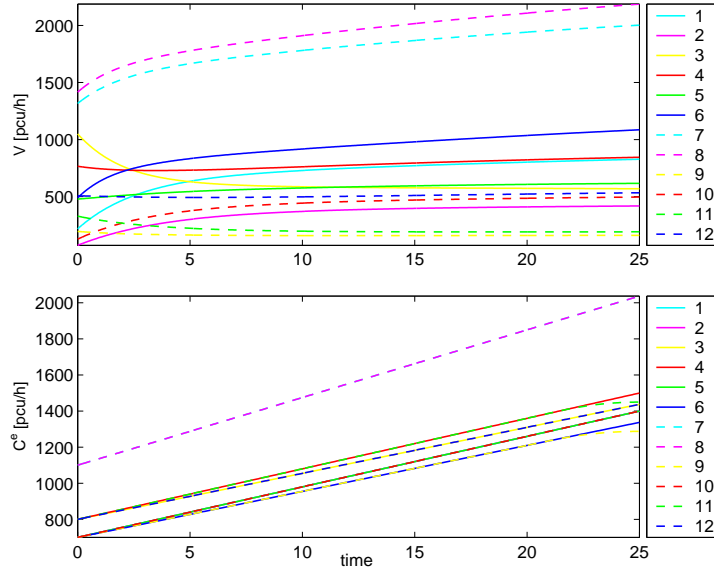


Figure 6.41: States $V_i(t)$ and $C_i^e(t)$ - case 3.2.

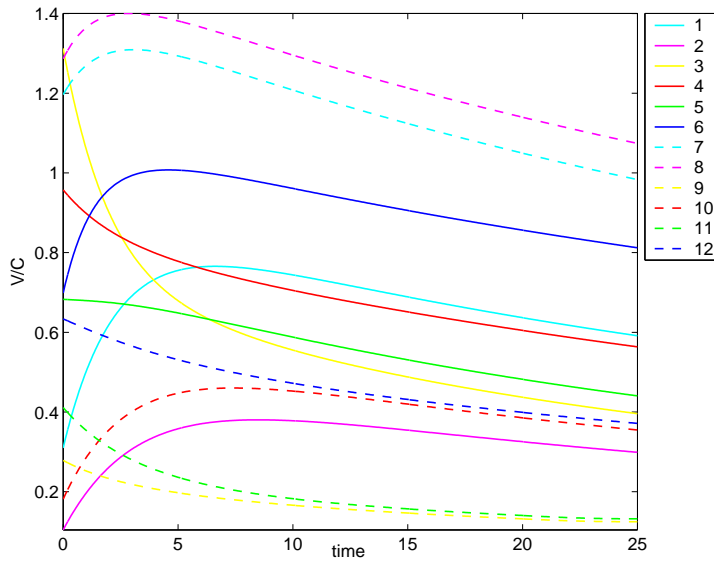


Figure 6.42: Level-of-service $V_i(t)/C_i^e(t)$ - case 3.2.

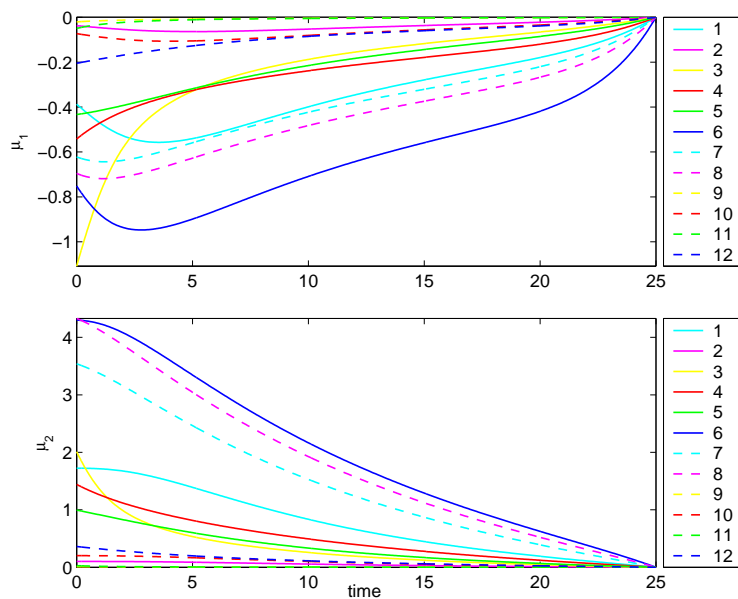


Figure 6.43: Costates $\mu_{1l}(t)$ and $\mu_{2l}(t)$ - case 3.2.

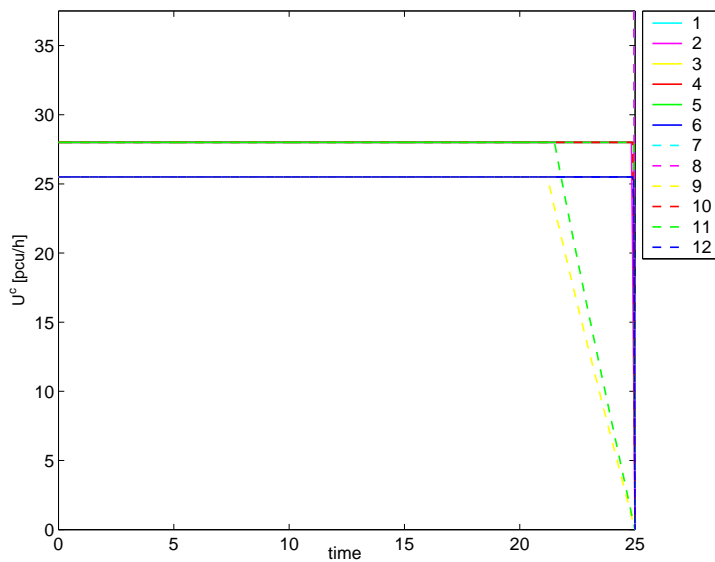


Figure 6.44: Control $U_i^c(t)$ - case 3.2.

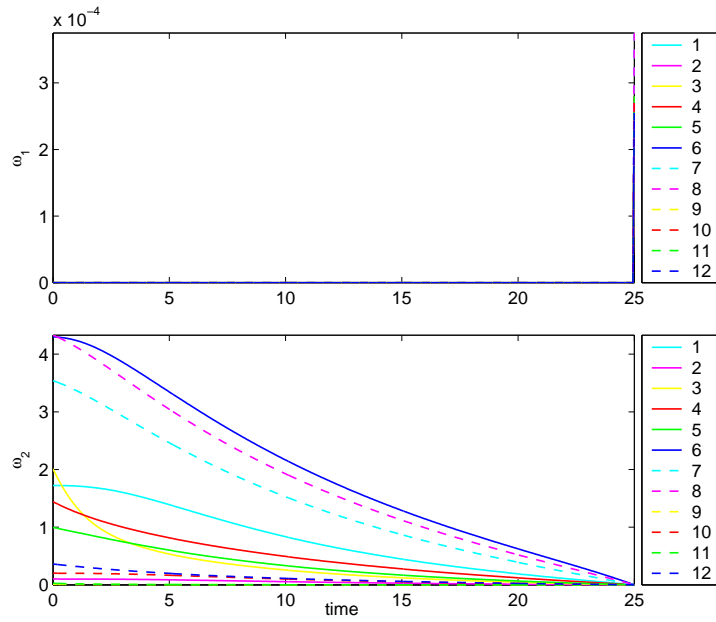


Figure 6.45: Multipliers $\omega_{1l}(t)$ and $\omega_{2l}(t)$ - case 3.2.

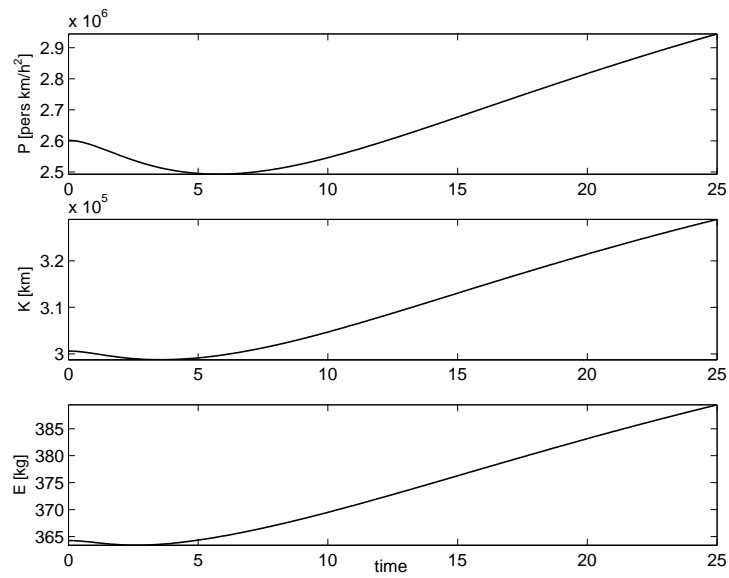


Figure 6.46: States $P(t)$, $K(t)$ and $E(t)$ - case 3.2.

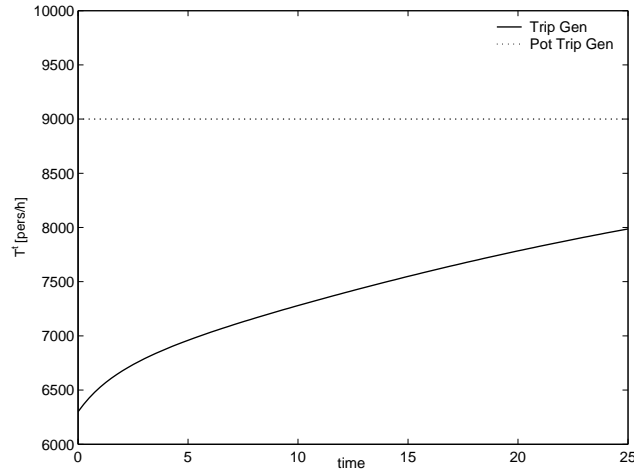


Figure 6.47: (Pot.) trip generation T^t - case 3.2.

maximum construction all the time, as there is no (state) constraint hindering this, hence forcing volume-over-capacity levels at all links to be monotonously decreasing over time. The total number of trips is indeed higher than in previous case study and reaches 7990 pers h^{-1} , which is about 400 pers h^{-1} more than in previous case study. Furthermore, the emission level is reaching 390 kg , compared to 365 kg in previous case study. In the next case study the emissions will be lower because of an emission constraint that is applied.

6.5.2 An optimal U^c -control with an integral emission state constraint for case 3.3 (A3/B1/C2)

In the eighth case study the previous case study has only been extended by the incorporation of an integral emission constraint to the integral emissions $\tilde{E}(t)$, similarly formulated as in equation (5.83), which is adjoined to the cost criterion as a penalty function.

Hence, the dynamic network travel demand model (A3) in combination with infrastructure supply model (B1), using cost criterion (C2) is:

$$\begin{aligned}
 & \max \int_{t_0}^{t_1} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \left(\tilde{Q}_i(t) \tilde{X}_j(t) \right)^{\frac{1}{2}} \exp(-\beta_9 c_{ij}^*(t)) \\
 & \quad - \sum_l \alpha_{71} U_l^c(t)^2 - \chi_1 \max \left(0, (\tilde{E}(t) - \tilde{E}_{t_1}^*) \right)^2 dt \\
 & \text{s.t. } \frac{dV_l(t)}{dt} = \gamma_1 \left(\hat{V}_l(t) - V_l(t) \right), \quad \forall l, \\
 & \quad \dots
 \end{aligned}$$

$$\begin{aligned}
& \dots \\
& \frac{dC_l^e(t)}{dt} = U_l^c(t), \quad \forall l, \\
& \frac{dE(t)}{dt} = \gamma_1 \left(\hat{E}(t) - E(t) \right), \\
& \frac{d\tilde{E}(t)}{dt} = E(t), \\
& \text{and } 0 \leq U_l^c(t) \leq U_l^{c\max}, \quad \forall l, \\
& \text{boundary conditions at } t_0 \text{ and } t_1,
\end{aligned} \tag{6.47}$$

where the disequilibrium demand $\hat{V}_l(t)$ is elastic as in equation (5.22), while the utility is a linear function of route travel time $\tau_r(t)$ only. The total integral emissions constraint for $\tilde{E}(t)$ is specified as in equation (5.83):

$$\tilde{E}(t) \equiv \int_{t_0}^t E(t') dt' \leq \tilde{E}_t^*, \tag{6.48}$$

and is added to the system as two coupled equations of motion for emissions, hence constraining total emissions over time. The first one is the standard disequilibrium emission model that calculates the actual state value, whereas the second equation is accumulating emissions over time, following equation (5.84) with boundary conditions as specified in equation (5.85). The upper-bound is added through the penalty function on total emissions $\tilde{E}(t)$ with: $\chi_1 = 0.03$. The environmental capacity $\tilde{E}_{t_1}^*$ is set at 9600kg, knowing that in the previous non-constrained case the total cumulative emissions were 9710kg at 7990pers h^{-1} . Furthermore, a smoothing term for the control $U_l^c(t)$, with: $\alpha_7 = 1.0 \times 10^{-6}$, is adjoined to the cost criterion. The partial derivative expressions for the functional J in this system are given in appendix D.2.

Given the penalty function on the total emissions, a different control path than in the previous case study should be observed. The origin-destination combinations with a high relative attraction value will be given preference when allocating the capacity enhancement resources. Because of this, a smaller number of total trips in the system can be expected.

The simulation results are depicted in figures 6.48 to 6.54. Indeed different control paths can be observed in figure 6.51, with some links keeping their original capacity, while others face a maximum value of new capacity added throughout the simulation, implying relatively high μ_{2l} values in figure 6.50. These links together form the important corridors in the network. The volume-over-capacity ratios for all links tend to decrease in figure 6.49, although not as fast as in the previous case study. Furthermore, it can be seen in figure 6.53 that the integral emission constraint is just binding at the end, at a level of 9600kg of volatile organic compound. The total trip generation has decreased to: 7910pers h^{-1} , as seen in figure 6.54. This reduction in trips seems rather

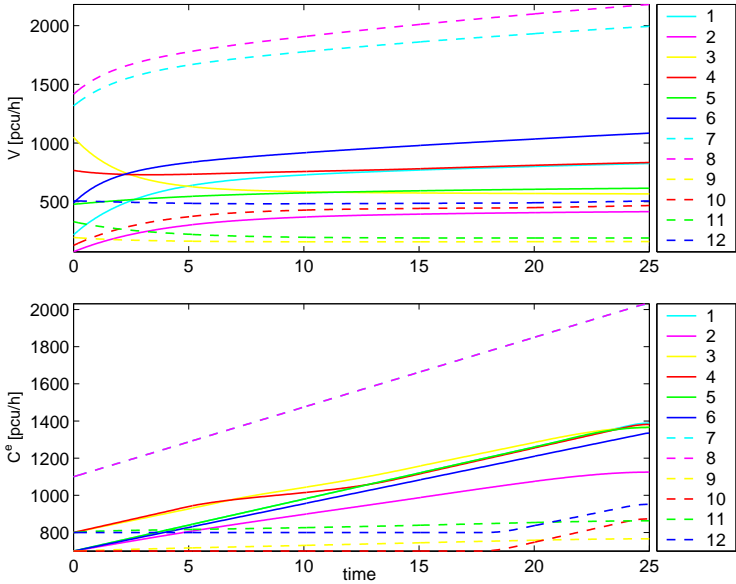


Figure 6.48: States $V_i(t)$ and $C_i^e(t)$ - case 3.3.

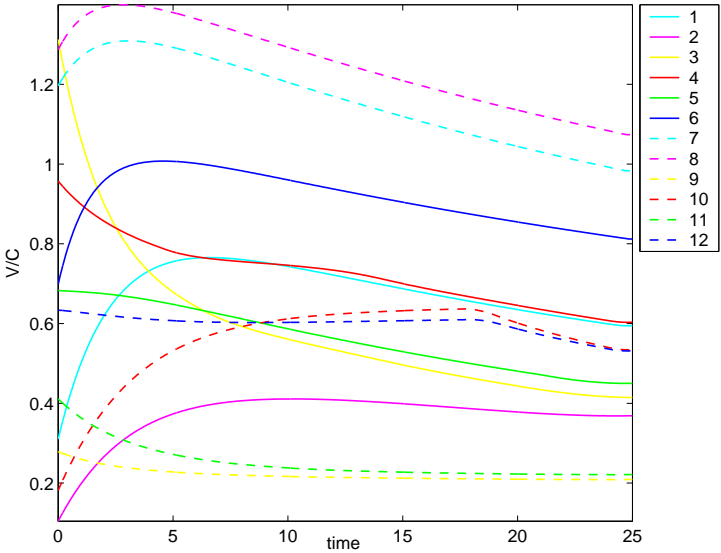


Figure 6.49: Level-of-service $V_i(t)/C_i^e(t)$ - case 3.3.

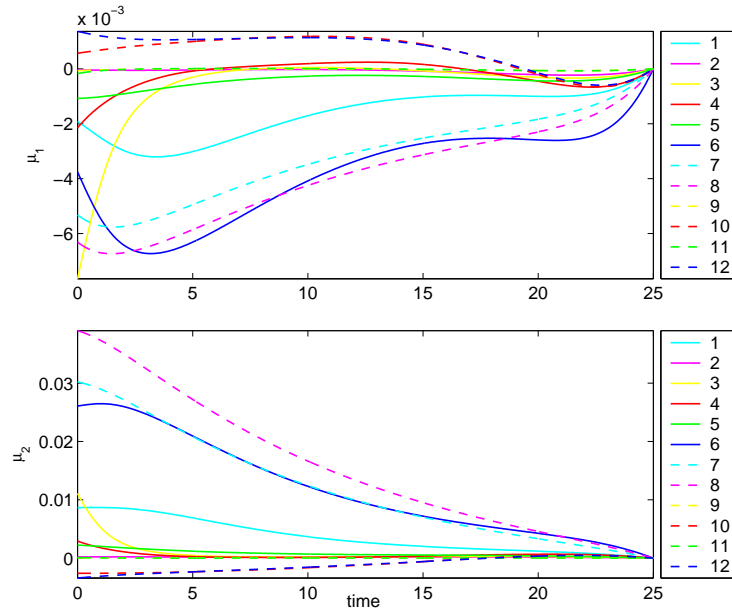


Figure 6.50: Costates $\mu_{1l}(t)$ and $\mu_{2l}(t)$ - case 3.3.

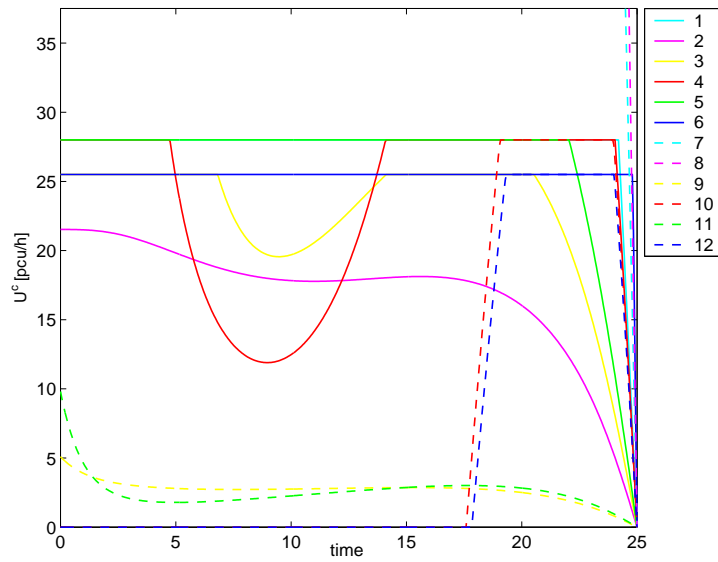


Figure 6.51: Control $U_i^c(t)$ (Boundary Value Problem) - case 3.3.

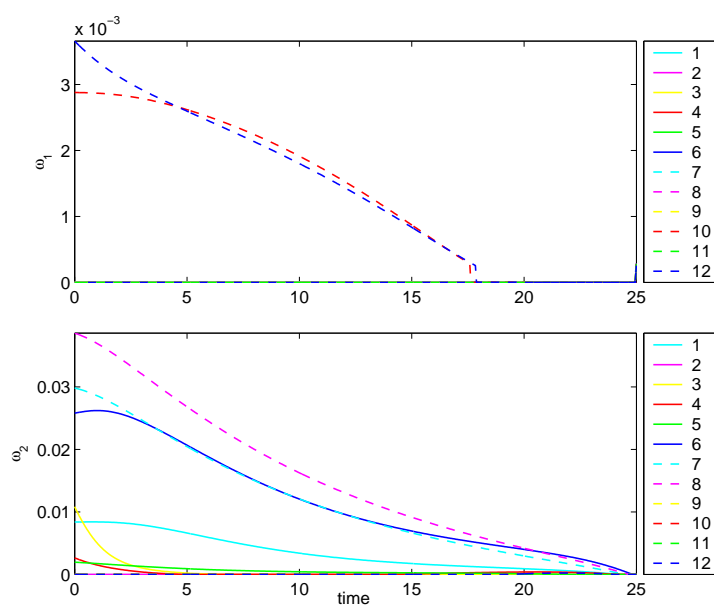


Figure 6.52: Multipliers $\omega_{1l}(t)$ and $\omega_{2l}(t)$ - case 3.3.

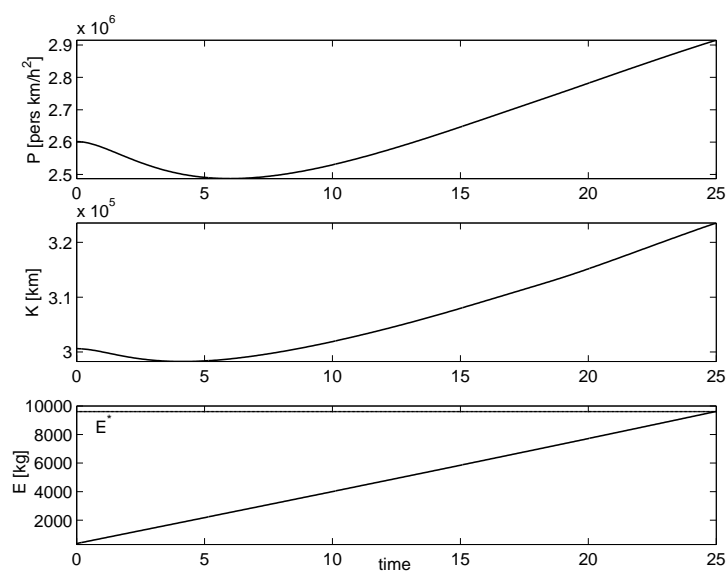


Figure 6.53: States $P(t)$, $K(t)$ and $\tilde{E}(t)$ - case 3.3.

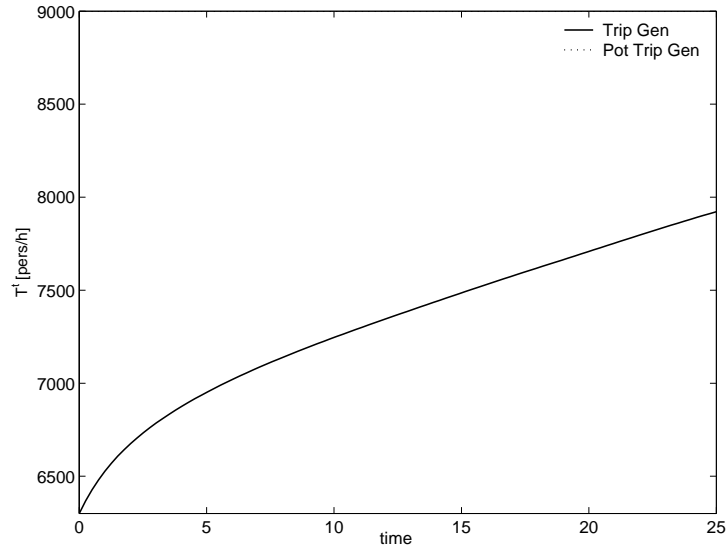


Figure 6.54: (Pot.) trip generation T^t - case 3.3.

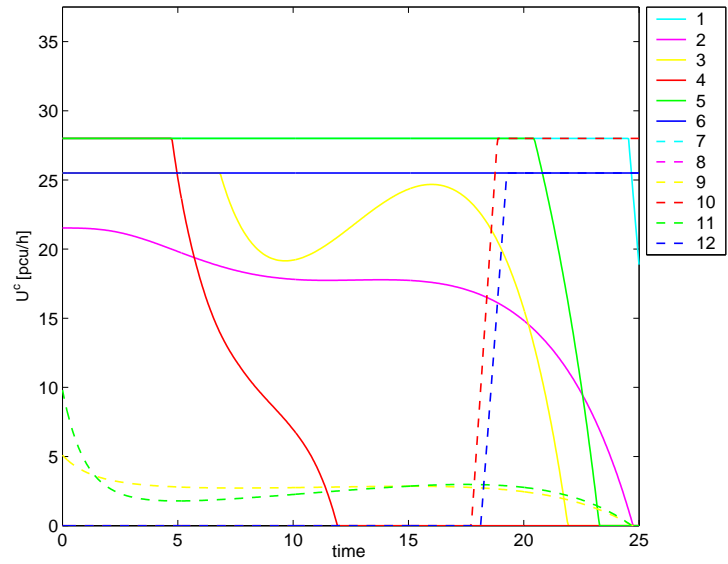


Figure 6.55: Control $U_i^c(t)$ (Initial Value Problem) - case 3.3.

low compared to the decrease in total emissions. The reason for this is that with dedicated controls to those corridors where a speed increase is most effective, large gains in emission reduction can be achieved, at a relatively high number of revealed trips.

Unfortunately, for this simulation a small residual: $r(t) = 14.983$, remains, which is most likely caused by the switching of the control variables as seen in figure 6.51. Because of this, an extra test on plausibility and robustness of this result needed to be performed. This is done by running an *initial value problem* simulation using the outcomes of the boundary value problem simulations. In other words, not the initial guess, but the calculated initial values $\mu_{1l}(t_0)$ and $\mu_{2l}(t_0)$ in combination with the initial values for the state variables, form the input to the initial value problem (IVP), which is accordingly solved by applying a variable-step continuous solver based on a Runge-Kutta (4,5) formula in MATLAB. Doing so, the optimal control paths in figure 6.55 remain. From this figure one can observe that the control paths have slightly changed, but not dramatically. Perhaps, only the control path for link 4 can be considered unreliable.

6.6 Case study: person throughput maximisation (C3)

The third cost criterion, which is considered in this chapter, is the one of person throughput maximisation. Remember that this cost criterion refers to maximisation of both quality of movement in terms of speed of person movement and quantity being moved in terms of person trips in the network. The cost criterion is applied to the small network only, which is figure 6.3, where link number 1 is now *exclusive* for public transport use (say mini-buses, mode $m=2$). The other links are exclusive for car use.

Furthermore, in this case study, as is the case in the next one, see paragraph 6.7 on equity maximisation, the exogenous variables $Q_i(t)$ will not be constant anymore as before (see for example table 6.5) but slightly vary in time, for instance resembling an increase in population over time in the origin zone. To do so, the systems have been rewritten as an autonomous systems, following equations (5.91) to (5.93). Obviously, it is also necessary to distinguish the set of modes $m \in \mathcal{M}$. Furthermore, the set of population segments $k \in \mathcal{K}$ is used, which is stable over time. In other words, there is no interchanging between population segments over time.

For this cost criterion only one case study is discussed. Parameter values differ a bit from previous case studies, to reflect different preferences of population segments, as well as different characteristics of the modes, which are considered.

6.6.1 An optimal U^t -control with an emission state constraint for case 2.5 (A4/B2/C3)

For the ninth case study, the dynamic network travel demand model (A4) in combination with infrastructure supply model (B2), using cost criterion (C3) is considered. By maximising the cost criterion, the productive capacity C^P of the network, which is equation (3.7), while complying with the emission state constraint, which is equation (5.82), can be determined. In addition, a quadratic smoothing term is adjoined to the cost criterion. The following system hence remains:

$$\begin{aligned}
& \max \int_{t_0}^{t_1} P(t) - \sum_l \alpha_{71} U_l^t(t)^2 - \chi_1 \max(0, (E(t) - E^*)) dt \\
& \text{s.t. } \frac{dV_l(t)}{dt} = \gamma_1 \left(\hat{V}_l(t) - V_l(t) \right), \quad \forall l, \\
& \quad \frac{dC_l^e(t)}{dt} = U_l^t(t) - \xi_1 C_l^e(t), \quad \forall l, \\
& \quad \frac{dP(t)}{dt} = \gamma_1 \left(\hat{P}(t) - P(t) \right), \\
& \quad \frac{dE(t)}{dt} = \gamma_1 \left(\hat{E}(t) - E(t) \right), \\
& \text{and } 0 \leq U_l^t(t) \leq U_l^{t\max}, \quad \forall l, \\
& \quad \text{boundary conditions at } t_0 \text{ and } t_1.
\end{aligned} \tag{6.49}$$

In order to allow for different population segments k , as in equation (5.71), the elastic disequilibrium travel demand $\hat{V}_l(t)$ now reads:

$$\hat{V}_l(t) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^l} \sum_{k \in \mathcal{K}} \theta_{1m} \left(\check{Q}_{i|k}(t) D_{ij|k}(t) G_{ijmr|k}(t) \right), \quad \forall l. \tag{6.50}$$

In addition, the utility function (5.26) is changed here to make it valid for multiple population segments k , as well as linear in route travel time only:

$$u_{ijmr|k}(t) = \beta_{3m|k} \tau_r(t), \quad \forall i, j, m, r, k. \tag{6.51}$$

Furthermore, the natural pavement deterioration rate: $\xi_1 = 0.01$, whereas the smoothing parameter: $\alpha_7 = 6.0 \times 10^{-5}$. The penalty function parameter: $\chi_1 = 1.3 \times 10^5$, while the environmental capacity is set at: $E^* = 48 \text{ kg}$. Assuming two population segments and two modes, the parameter values depicted in table 6.7 hold.

Similar to the unconstrained accessibility maximisation case study, it is expected that a control path with maximum values for capacity enhancement, until

Table 6.7: Mode and population segment specific parameter values.

$\beta_{3m k}$	$k=1$	$k=2$	θ_{1m}
$m=1$ (car)	1.50	2.00	1.0
$m=2$ (bus)	1.00	1.25	0.1
%k	0.35	0.65	-

the environmental capacity constraint becomes active, is revealed. The maximum person throughput level that is obtained, is then the productive capacity C^P of the network.

The simulation results are depicted in figures 6.56 to 6.61. Indeed the control paths are at their upper-bound part of the time as can be seen in figures 6.57 and 6.59. The bang-bang structure is very well visible. The smoothing part of the control path seems absent, mainly because the costate paths, in figure 6.58, are very steep and descend from values in the order of 1.0×10^4 towards zero. In figure 6.56 the state trajectories are shown. The volume (in [pcuh⁻¹]) on link 1 is very low, but increases over time as more trips are being made by mini-bus. The number of person trips by mini-bus, hence increases, while the travel time decreases because of continuous capacity enhancement. The gain in number of passenger trips (the objective in this case study) is best made through link 1 as can be seen from the costate trajectories for link 1 (relatively large negative μ_{11} and relatively large positive μ_{21}), in figure 6.58, as well. The volume-over-capacity levels, in figure 6.57, can be seen to decrease most of the simulation, but increases as soon as capacity enhancement is brought to a stop. The productive capacity of the network can be found in figure 6.60, which is: $C^P = 9.6 \times 10^5$ perskmh⁻². Furthermore, a small exceeding of the environmental capacity constraint to a level of: $E(t) = 49.0$ kg, similar to figure 6.31, can be observed. The main reason for this is probably that it is much harder for the solver to work with this particular cost criterion than with the others discussed in this chapter. The other three cost criterions are nonlinear (actually quadratic or negative-exponential), whereas this one is linear, hence no explicit direction giving derivative remains. The variability in social-economic variables for the origin zone can be best seen in figure 6.61, where the jump in potential trip generation $\tilde{Q}_i(t)$ is clearly visible. The total trip generation is about 2445pers h⁻¹. Without environmental constraint the emission level would have been: $E(t) = 51.1$ kg, at a total trip generation rate of 2490pers h⁻¹.

This case study would, ideally, have been combined with cost controls, for example for vehicle tax $U^v(t)$ and parking tax $U_j^p(t)$ as discussed in equation (5.28). It is much easier, common and logically, to reflect individual preferences through such monetary parameters instead of value-of-time parameters as used here. However, singularity problems prevent inclusion of these controls for the moment. See also paragraph 6.4.4 in this chapter.

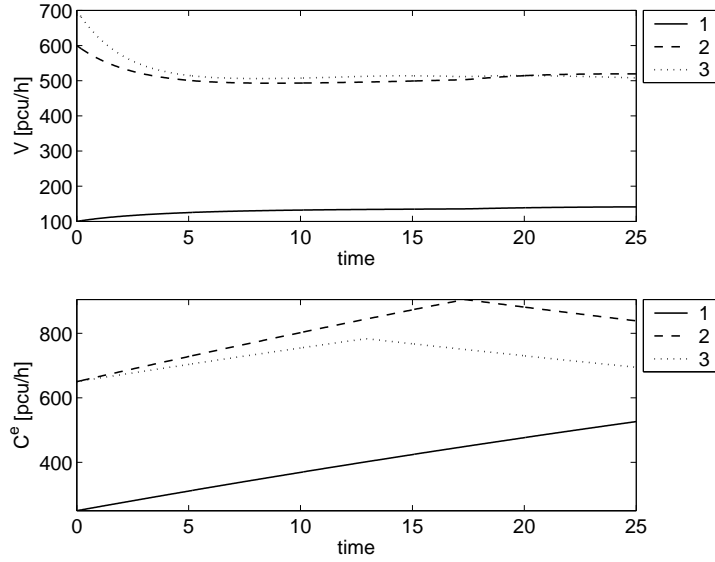


Figure 6.56: States $V_i(t)$ and $C_i^e(t)$ - case 2.5.

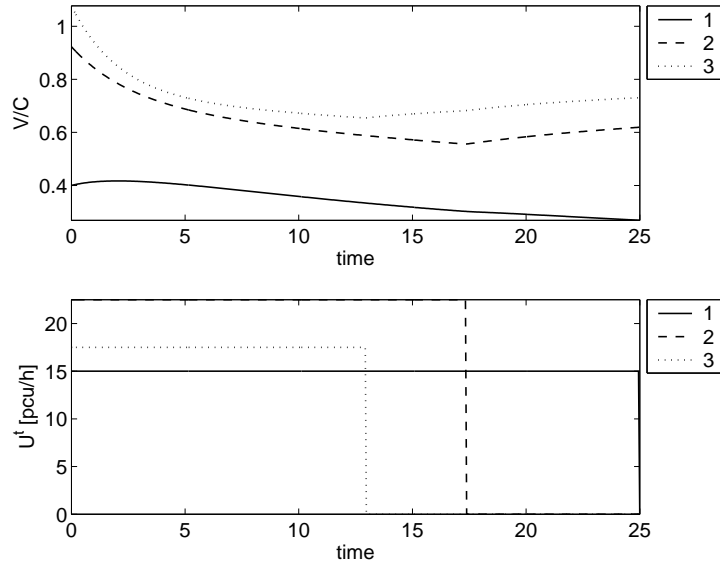


Figure 6.57: Level-of-service $V_i(t)/C_i^e(t)$ and control $U_i^t(t)$ - case 2.5.

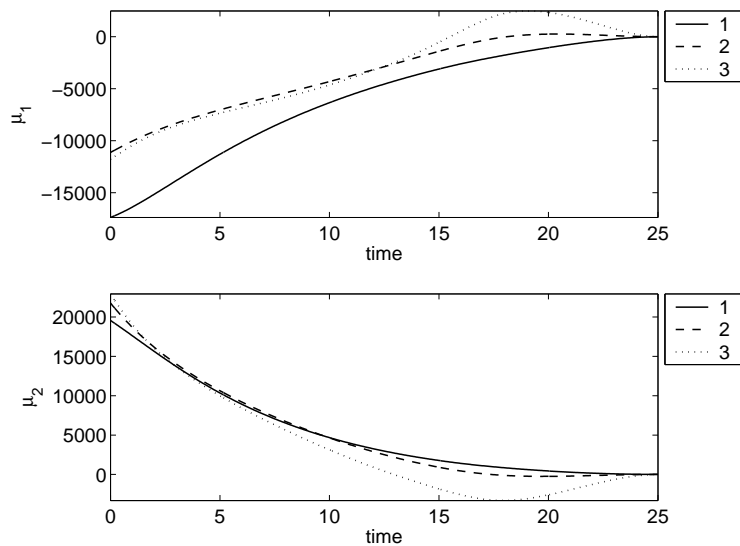


Figure 6.58: Costates $\mu_{1l}(t)$ and $\mu_{2l}(t)$ - case 2.5.

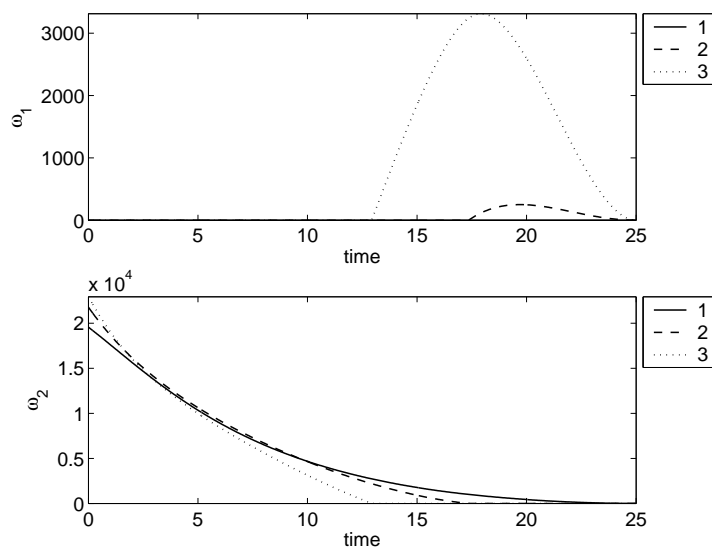


Figure 6.59: Multipliers $\omega_{1l}(t)$ and $\omega_{2l}(t)$ - case 2.5.

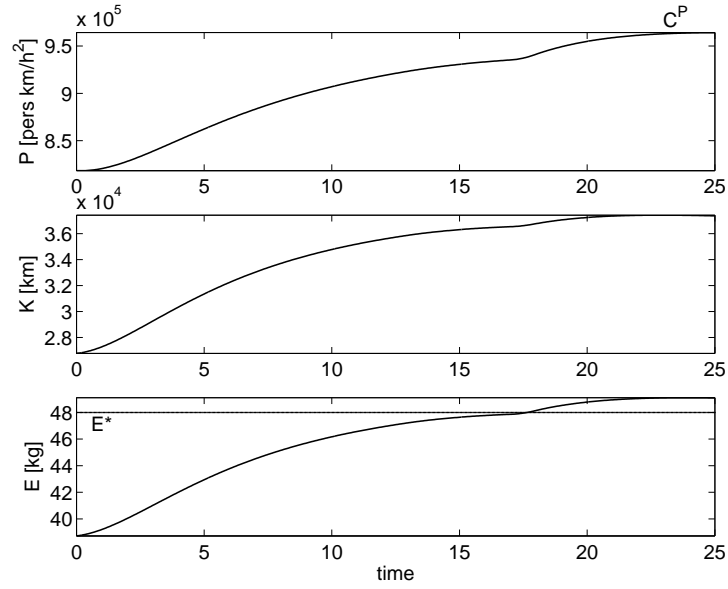


Figure 6.60: States $P(t)$, $K(t)$ and $E(t)$ - case 2.5.

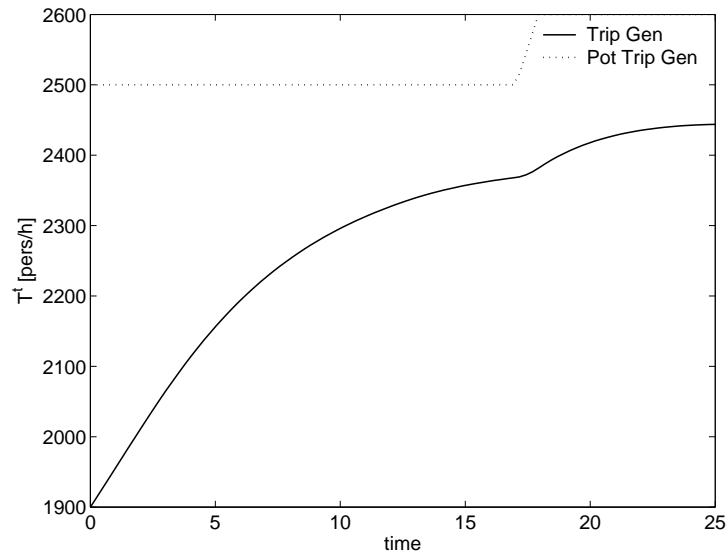


Figure 6.61: (Pot.) trip generation T^t - case 2.5.

6.7 Case study: equity maximisation (C4)

The fourth and final cost criterion, which is considered in this chapter, is the one of equity maximisation. Remember that this cost criterion, the way it is defined here, refers to allocating resources in such a way that the average composite costs experienced by all population segments (with their individual preferences) is deviating minimally from those originally experienced (thus at time: $t = t_0$). Once more, it should be mentioned that this formulation is not reflecting the original definition of equity as posed in paragraph 3.2.4, but rather a definition that comes closest to this, while keeping track of the mathematical requirements for an appropriate cost criterion. The cost criterion is again applied to the small network only, which is the network in figure 6.3, where link number 1 is, again, exclusive for public transport use (mini-buses, mode $m=2$). The other links are exclusive for car use. Also the exogenous variable $Q_i(t)$ varies with time.

For this case study it is also necessary to introduce a set of population segments $k \in \mathcal{K}$, as well as to distinguish a set of modes $m \in \mathcal{M}$. Hence, the parameter values in table 6.7 still comply.

6.7.1 An optimal U^t -control with an emission state constraint for case 2.6 (A4/B2/C4)

For the tenth and final case study the dynamic network travel demand model (A4) in combination with infrastructure supply model (B2), using cost criterion (C4) is considered. Furthermore, an emission state constraint, alike equation (5.82), is applied. In addition, a quadratic smoothing term is adjoined to the cost criterion. The following system hence remains:

$$\begin{aligned}
& \max \int_{t_0}^{t_1} - \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \left(\sum_{k \in \mathcal{K}} \check{Q}_{i|k}(t) c_{ij|k}^*(t) - \bar{c}_{ij}^*(t_0) \right)^2 \\
& \quad - \sum_l \alpha_{\gamma_1} U_l^t(t)^2 - \chi_1 \max(0, (E(t) - E^*))^2 dt \\
& \text{s.t. } \frac{dV_l(t)}{dt} = \gamma_1 \left(\hat{V}_l(t) - V_l(t) \right), \quad \forall l, \\
& \quad \frac{dC_l^e(t)}{dt} = U_l^t(t) - \xi_1 C_l^e(t), \quad \forall l, \\
& \quad \frac{dE(t)}{dt} = \gamma_1 \left(\hat{E}(t) - E(t) \right), \\
& \text{and } 0 \leq U_l^t(t) \leq U_l^{t \max}, \quad \forall l, \\
& \quad \text{boundary conditions at } t_0 \text{ and } t_1,
\end{aligned} \tag{6.52}$$

where $\hat{V}_l(t)$ is the elastic disequilibrium travel demand model from equation (6.50). Furthermore, the natural pavement deterioration rate: $\xi_1 = 0.01$,

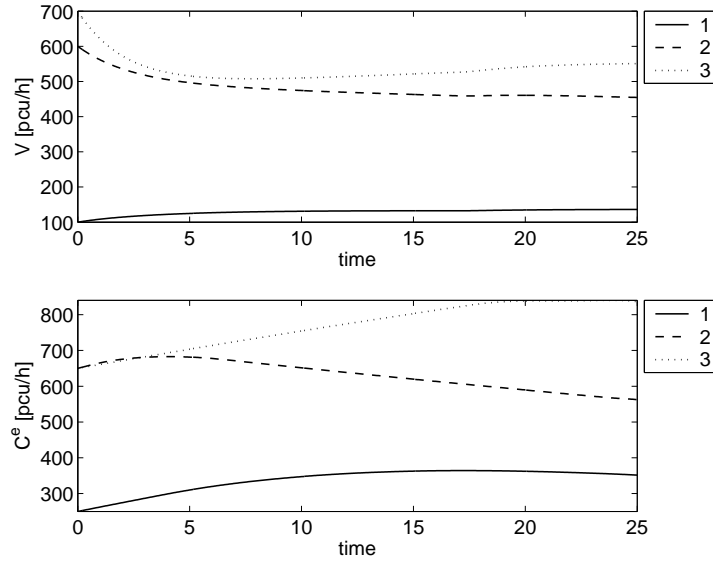


Figure 6.62: States $V_i(t)$ and $C_i^e(t)$ - case 2.6.

whereas the smoothing parameter value is: $\alpha_7 = 6.0 \times 10^{-5}$. The penalty function parameter: $\chi_1 = 0.75$, while the environmental capacity is set at: $E^* = 48 \text{ kg}$.

For this case study one would expect control paths for capacity enhancement to be at their maximum value for some part of the time for all links. From table 6.2, it may be concluded that persons from population segment: $k = 2$, which happens to be the larger group in this case study, show higher sensitivity to an increase in travel time, in particular for mode: $m = 1$, which is the car. Hence, using this formulation, the links where you will have most car users from segmentation: $k = 2$, you expect to see attempts to keep track with induced demand through capacity enhancement, more than on other links.

The simulation results are depicted in figures 6.62 to 6.67. The control paths can be seen in figure 6.63 and show that road link 3 is given much more allocation of resources than in the previous case study (with exactly the same input data), while links 1 and 2 only need marginal investments to keep track of the initial composite costs on that corridor, as can also be observed in figure 6.64. The smoothing part of the control path is much more visible as well. In figure 6.62 the state trajectories are shown. The volume (in $[\text{pcu h}^{-1}]$) on link 1 is very low, but increases over time as more trips are being made by mini-bus. The number of person trips by mini-bus, hence increases, while the travel time decreases because of continuous capacity enhancement. The gain in number of passenger trips is slightly less than in previous case study where more capacity was added to the exclusive bus link. The volume-over-capacity levels in

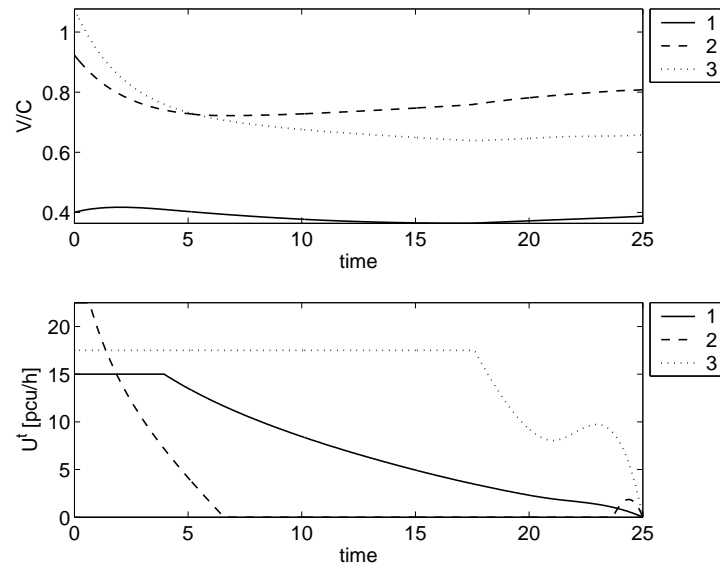


Figure 6.63: Level-of-service $V_i(t)/C_i^e(t)$ and control $U_i^t(t)$ - case 2.6.

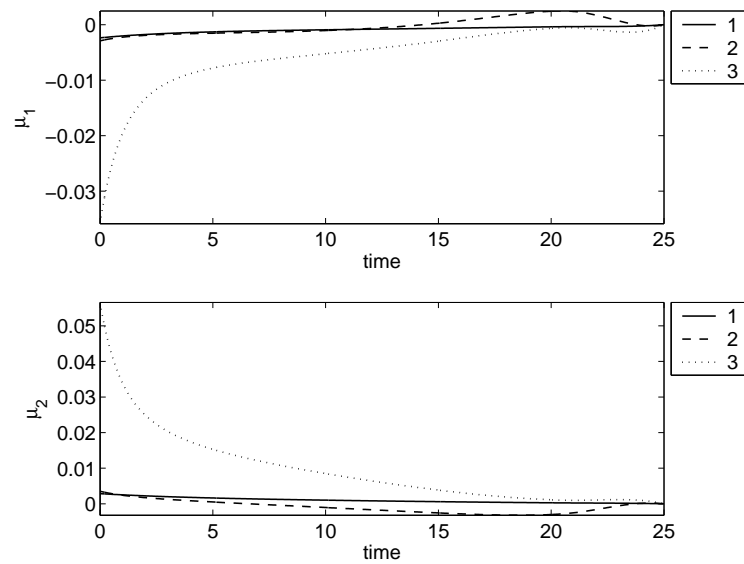


Figure 6.64: Costates $\mu_{1l}(t)$ and $\mu_{2l}(t)$ - case 2.6.

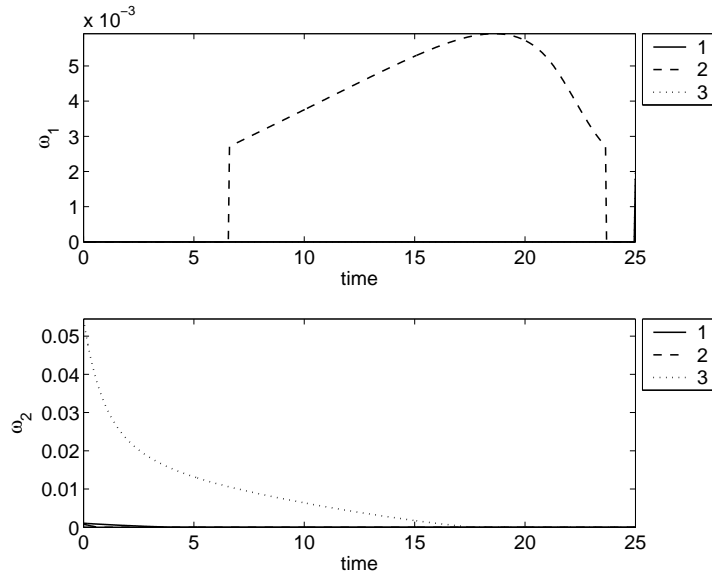


Figure 6.65: Multipliers $\omega_{1l}(t)$ and $\omega_{2l}(t)$ - case 2.6.

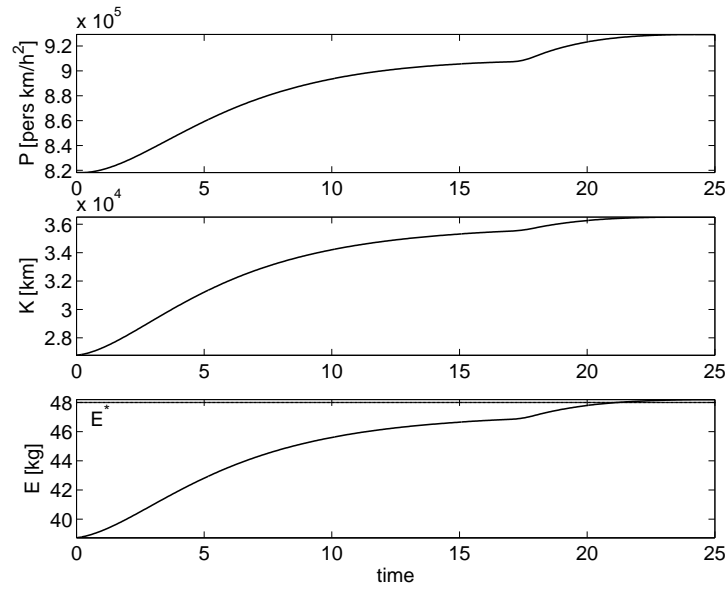


Figure 6.66: States $P(t)$, $K(t)$ and $E(t)$ - case 2.6.

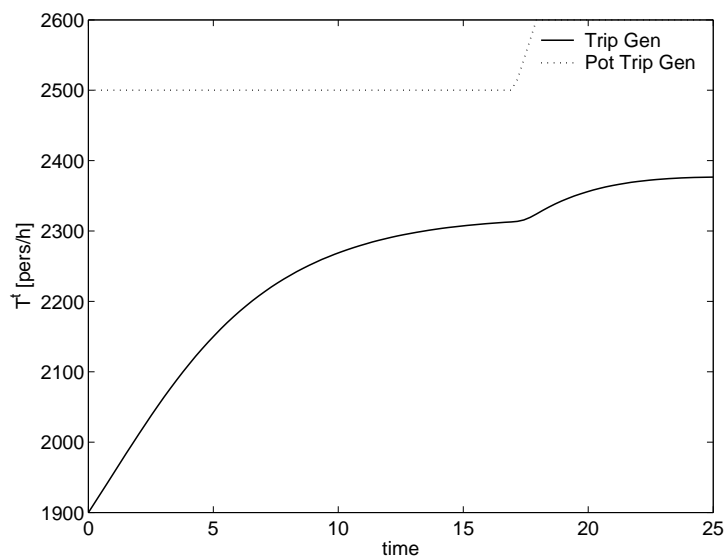


Figure 6.67: (Pot.) trip generation T^t - case 2.6.

figure 6.63 can be seen to decrease most of the time, but increases as soon as capacity enhancement is brought to a stop. In figure 6.66, it can be observed that the environmental capacity constraint is just binding at the end of the simulation at: $E(t) = 48.2\text{kg}$. The variability in social-economic variables for the origin zone can be best seen in figure 6.67, where the jump in potential trip generation $\tilde{Q}_i(t)$ is clearly visible, similar to previous case study. The total trip generation is about 2375trips h^{-1} . Without environmental constraint the emission level would have been: $E(t) = 50.0\text{kg}$, at a total trip generation rate of 2460trips h^{-1} .

6.8 Opportunities for a real-life application

Previous case studies have demonstrated the applicability of the dynamic transport model for different type of road networks under various behavioural assumptions. Even though the computational burden of solving the different models appeared to be significant, the results show that a real-life application should be possible, provided some refinements in the model equations, algorithms and coding are accomplished (as will be recommended in chapter 7). However, in spite of any refinement, it is not likely that it can be applied directly to the size and scope of modern-day transport networks, which may easily have a network topology of thousands of nodes, ten thousands of links and hundreds of origin/destination pairs, not even mentioning the number of user classes and vehicle types that are considered. Therefore, this paragraph looks ahead to such real-life application.

The intended use of the dynamic transport model presented in this chapter has, in paragraph 1.2.3, been typified as comprehensive to strategic, implying a medium to long time decision horizon in combination with a relative low detail level of analysis. At this level the optimal directions in which a sustainable transport policy should be developed are identified, taking into account future social-economic developments in the area of study. These policies are based on the model equations set forward in chapter 5 and the case studies presented in this chapter. The case studies are illustrative for the type of questions that can be, and cannot be, answered using the strategic model. Most striking is that the control paths basically encompass changes in general road link capacity, whereas the intended use of the financial control variables is of a similar broad and strategic level. Hence, these control paths have to be translated into operational measures, like those mentioned in paragraph 3.4.1. In addition, continuous capacity control paths will have to be translated into discrete units of road capacity if these are to be road lanes at some point.

The features in the presented model, which justify the term ‘strategic’ are, besides the intended time horizon of decision-making, first and foremost, the limited number of zones and links that can be modelled, as well as the general meaning of the control variables. In the case studies the number of zones and links have, on purpose, be kept low, although it should not be a problem to increase this. However, this number will always have to be in the order of, say 10 to 20 zones, with about 50 to 100 links, which should be enough for a typical strategic case study. Above this number, computation times get too long and the manual treatment necessary for setting initial guesses as well as parameters like the smoothing parameter α_7 and the penalty parameter χ_1 , becomes very difficult as they have to be in the right order of magnitude (which is not exactly known in advance). The intended use of the strategic dynamic transport model is therefore of greatest value when combined with another transport model at a more operational level, which may be static. A very interesting example is provided in Al-Azzawi (2001), who, using the static transport model EMME/2, combined two levels of analysis in a hierarchical modelling exercise.

To do so for the dynamic model presented here, the model should be tailored to link with a detailed and operational level:

1. upward from a detailed network model for the generation and update of strategic supply data (regarding zones and network) and sustainable transport strategies;
2. downward to a detailed network model to allow the detailed implications of network changes and related forecast changes in trip generation to be made more detailed as well as tested.

This requires an hierarchical modelling framework with an upper-level model with an aggregated network supply and aggregated zonal representation, in combination with a lower-level detailed network performance, as illustrated in figure 6.68. Such detailed model could be a static traditional travel demand

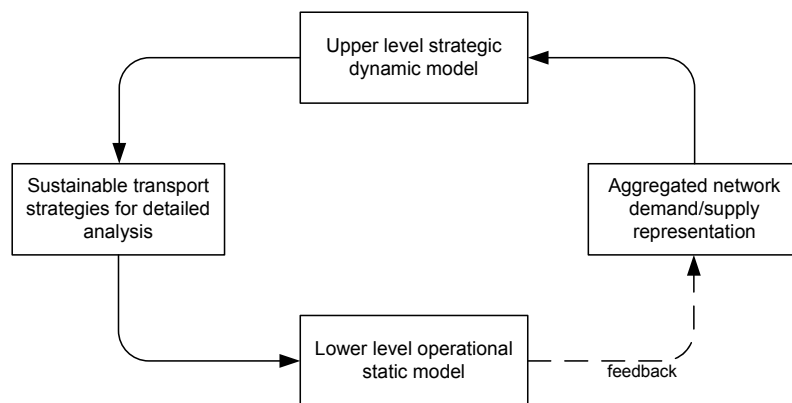


Figure 6.68: Hierarchical modelling approach

model, but also a dynamic¹¹ microscopic model. Both type of models are best deployed at a number of points in time (at least at the endpoint in time) and could provide information for refinement or detailing in the zonal and network aggregation itself, hence providing a feedback, back to the upper-level, again.

Some practical methods, as for example presented in Al-Azzawi (2001), and fine methods for zonal aggregation as well as network aggregation exist and need to be applied in that case. For zonal aggregation, it is important to take good consideration of the spatially detailed trip generation data that exists. In particular, the treatment of intra-zonal traffic becomes an issue here, as localised variations within the aggregated zone gets lost. An early introduction on the effects of zonal aggregation as well as network detail on transport model performance is provided in Jansen and Bovy (1982). For aggregated transport networks the spatially aggregated model flows and travel times are not on individual links but on aggregated links, representing corridor capacity. A good overview on network aggregation is given in Friesz (1985), as well as a recent operationalisation in Van Gent (2005).

¹¹Dynamic in the sense of short-term dynamics only.

Chapter 7

Summary, conclusions and further research

7.1 Summary

Current developments in urban transport realities force authorities to plan, manage and maintain their transport systems more accurately and take into account the requirements of a growing number of complex and sometimes conflicting interests like congestion relief, pollution reduction, efficient resource use, equity and accessibility.

The common solution to these emerging problems and changing requirements in transport is to build extra capacity, make better use of existing infrastructure, discourage and/or promote other means of transport or even influence travel patterns of people as well as freight, following the principle of *predict-provide-manage*. Decisions to do so are supported by best-practices, theories, and tools of transport planning. However, by doing so, current transport systems and transport planning methods and models (used in developed as well as developing countries) are not necessarily compatible with the requirements of sustainable transport development. Adequate transport systems can only be obtained by use of a new sustainable transport paradigm and accompanying analytical framework. Therefore, this thesis has presented a theoretical framework, together with a methodology to better incorporate requirements of sustainable development into models for transport policy and planning, in particular travel demand models. With the knowledge and outcomes presented in this thesis, it should become possible to make more effective and efficient use of available and affordable scarce resources for enhancing transport system performance.

The theoretical framework is based on a paradigm for sustainable urban transport development, which is developed in this research. This paradigm advocates a more efficient, equitable, and environmentally sensitive transport system, in-

spired by the global message on the need for sustainability by leading scientists and politicians, who demonstrated that unchecked growth on our planet may lead towards environmental and ecological 'overshoot' and pending disaster. A sustainable transport development paradigm goes further and advocates a comprehensive decision-making that anticipates and manages scarce resource use, including environment and finance, while developing the transport system in terms of quality of access and/or person throughput.

To aid in the complex process of transport policy and planning, during the past few decades travel demand modelling tools have evolved from a simple heuristic towards an integrated four-step sequential model with advanced sub models for trip generation, trip distribution, modal choice and traffic assignment. However, the four-step model is, particularly in view of the sustainable transport development paradigm, being criticised for several reasons, amongst others because of the sequential and static nature, hence lacking feedback options. Furthermore, because of the inelasticity of the trip generation sub model to accessibility, hence denying the existence of generated travel demand. In addition, for the typical predict-provide-manage approach in using these models, which can be classified as demand-driven, whereas resource limits demand a supply-limited, or provide-manage-predict approach, which offers the possibility to calculate an optimal package of traffic and transport related measures that go with an explicitised transport policy objective and given resource capacities.

Besides this, it is claimed that conventional transport models can hardly produce useful recommendations to decision-makers if they are not founded on the understanding of the continuously changing behaviour of its users (the trip makers), the performance of the transport system itself as well as the complex and interacting objectives by its decision-makers. Not the individual decisions of these actors (both the decision-makers as well as users) should guide actions, but only the coevolving state of their individual systems should do so. Dynamic optimisation that is able to cope with these critiques, while producing a prescribed sequence and timing of actions to meet specific transport objectives by a given time, is then a logical evolutionary step in transport modelling.

Such a dynamic model that can assist in this complex and political decision-making process with respect to sustainable transport development, has been introduced and developed here, based on the conceptualisation and characterisation of the sustainable urban transport development problem as an optimisation problem. The basic building blocks of the traditional four-step modelling by itself, are believed to be adequate for the questions at hand, and have thus formed the basis for the dynamic modelling. A proactive approach that can be characterised as one of provide - manage - predict (hence 'prevent'), in contrast to the predict - provide - manage approach, then results.

The dynamic model adopts a disequilibrium formulation that allows for partial-adjustment of state variables, because of time-lags to the desired equilibrium level. The model features common transport state variables like travel demand

(per link in the network), infrastructure supply (effective link capacity), kilometres travelled in the network, person throughput in the network, and total emissions of pollutants in the network, which are all modelled using ordinary differential equations. Travel demand, derived from social-economic variables in the zones using a discrete-choice model, analogue to the gravity model, is also elastic to changes in transport system performance, whereas the pavement quality deteriorates autonomously or through wear caused by axle loading, which affects the effective capacity accordingly. As travel demand and infrastructure supply continuously search for their desired equilibrium, transport system performance changes because of induced travel, but also due to changes in destination, mode as well as route choice. Different control variables aiming at changing this equilibrium are at stake to the transport planner, including supply-side control variables like changing capacity and applying maintenance, but also pricing variables to affect the demand-side. Guided by different quantified transport policy objectives, like congestion minimisation, accessibility maximisation, person throughput maximisation or equity maximisation, as well as different type of constraints to the control values or to a state variable, for example total emissions, controls are effected in different magnitudes for different links, and at different moments in time.

The model in its present form can be applied directly to strategic networks of limited numbers of (aggregated) zones and (aggregated) links. Derived as a two-point boundary value problem in continuous-time, based on Pontryagin's Maximum Principle, the dynamic model reveals control paths for achieving a sustainable and developed transport system, while also providing information on the values for the state variables and their associated costate variables over time. The model has, however, not been implemented using a full set of real world data yet, hence no real policy implications from the example case studies can be drawn. The main reason for this being the apparent lack of robustness of the chosen solution method, in particular with respect to the interaction of control variables, but also because of the inclusion of nonlinear control variables. However, from the initial development of methods to solve the complicated model, i.e. the case study results, it may not be long before the full model can be implemented, estimated and applied in strategic transport planning.

Detailed conclusions and recommendations for further research are given in the remainder of this chapter.

7.2 Conclusions

In this paragraph, the main conclusions and reflections are presented. First, some conclusions are given that deal with the description of the sustainable urban transport development problem itself, followed by conclusions that are related to the formal characterisation of the sustainable transport development problem, the conceptual development as well as the accompanying require-

ments. In addition, conclusions with respect to the dynamic model and the chosen solution method are discussed, making a distinction between the model equations themselves, the applied optimisation method as well as the computational aspects. Furthermore, some conclusions related to the case studies are given. Finally, all conclusions and reflections accumulate in the answering of the research questions from chapter 1.

7.2.1 The urge for sustainable urban transport development

Mobility of people (and freight) is an essential prerequisite for social-economic development. In most cities throughout the developed and developing world, however, motorised vehicles, notably cars and trucks, have become the most important means of mobility, at the cost of non-motorised transport as well as public transport.

Because of this motorisation, congestion, traffic unsafety, air and noise pollution, changing land-use patterns, social isolation etceteras, have become common and widespread images in cities. These images form a serious threat to the wish for enhancing social-economic opportunities, enabled by this mobility in the first place. Hence, the attractiveness of cities is at stake, the number of transport-deprived (those who don't have access to transport or cannot afford transport) rises and health-damaging effects become a serious concern. Large investments in infrastructure supply and travel demand (management) are accordingly necessary, which amongst others put a large financial burden to available budgets.

Therefore, a vision for a sustainably developed transport system needed to be developed. One where person transport, accessibility, quality of life, environment, congestion, equity etceteras, have an important role, while at the same time taking care of the generations ahead in terms of financial and environmental capacities. Doing so, cities may again become attractive to live, work and reside in, now and in the future.

The general analytical framework for transport planning, i.e. the traditional transport planning model, which is primarily based on the theory of travel demand modelling that is frequently used nowadays (in developed as well as developing countries), is only partly useful for quantifying transport system performance and impacts with respect to issues of sustainable development and transport. This, because outcomes of these methods and models are not necessarily compatible with the requirements of sustainable transport development. It is claimed that adequate transport systems can only be obtained by use of the new sustainable transport paradigm and accompanying analytical framework. Hence, the need for a renewed analytical framework for transport planning, where requirements for sustainable development, in terms of environmental, financial and social considerations, are internalised, has been identified.

This analytical framework has as a basis the common aspects of the traditional transport planning model in deriving information for making decisions on the future development and management of transport systems, especially in urban areas. Apart from their proven usefulness in the decision-making process, these models are also known for having several shortcomings as lacking possibility to explicitly incorporate transport policy objectives, accessibility, land-use feedback, travel demand elasticity etceteras, especially when it comes to issues of sustainable development. The main shortcomings - within the scope of this thesis - related to incorporating requirements of sustainable development in the strategic transport planning process, in particular those of travel demand elasticity and accessibility, have thus been tackled in this research. The basic building blocks of the traditional four-step model by itself, however, are believed to be adequate for the questions at hand, and have thus formed the basis for the modelling.

7.2.2 Characterisation of the problem

Based on the underlying principle of sustainable development, the current threats and damages to health and environment make current transport realities unacceptable. Hence, in this research a sustainable and developed urban transport system has been postulated to be: ‘a transport system that meets the people’s transport related needs in terms of mobility, accessibility and safety, within limits of available or affordable environmental, financial and social resource capacities’.

This definition is based on the basic ideas of sustainable development and its taxonomy that characterises sustainable development as consisting of two distinct elements, that is sustainability as well as development, while having three different dimensions, that is economic and financial sustainability, environmental and ecological sustainability as well as social sustainability, furthermore having an analytical, a normative as well as a strategic level of discourse.

On the basis of this, a framework for sustainable urban transport development has been constructed that adopts a multi-directional conception of sustainable transport development, which tries to answer two important main questions, i.e.:

1. How can basic mobility and accessibility options to people be sustained or enhanced ? [the development question]
2. How can limited transport related resources, that is environmental, social and economic resources capacities, be used to guarantee intergenerational equity? [the sustainability question]

In addition, from an analytical point-of-view, sustainable transport development is regarded as a process, being intrinsically dynamic, as it is the coevolution of the individual subsystems that makes a sustainable transport development.

Applying this conceptual framework, transport professionals should now be able to aim directly at reaching certain transport development objectives through their transport policies and plans, while maintaining non-declining levels of transport system performance as well as keeping resource-use levels below those maximally acceptable, in other words affordable or available, levels.

To implement this in transport planning and modelling practice, several requirements have been derived and implemented, in terms of:

Policy objectives Transport system performance (over time), which is the quality of functioning of the transport system at a given level of travel demand and infrastructure supply, should be explicitly related to the political direction chosen, which is accordingly translated into a (quantifiable) transport policy objective and implemented directly into the modelling. These objectives of transport system performance may vary from a mere motorised traffic orientation, in terms of level-of-service, to social indicators, in terms of equity, or even (weighted) combinations of objectives;

Transport dynamics Travel behavioural rules and transport system mechanisms have been studied in transport science, and implemented in transport models, extensively. A distinction is usually made in mechanisms of travel demand and infrastructure supply, that are considered to be in an equilibrium. Travel demand is related to the social-economic realities of people (in terms of utility of trip making) as well as the transport network accessibility available to the people (in terms of disutility of trip making). Hence, travel demand is elastic to changes in both the social-economic realities as well as accessibility, and should be implemented in the modelling as such. Furthermore, these different mechanisms operate at different time-scales. A change in route choice-behaviour might require a relatively short period to transpire, whereas a change in infrastructure capacity might take a longer period to be accomplished. Therefore, it seems unlikely that an equilibrium between demand and supply exists in a dynamic model. A disequilibrium formulation is thus considered to be more appropriate. Hence, a system dynamics approach, including lagged-adjustment, disequilibrium models, should be used;

Resource capacities Quantitative targets, preferably related to international standards for resource-use, for example those of the WHO or the legally binding targets for greenhouse gas emissions of the Kyoto protocol, should be set and internalised in transport models. To do this within the framework derived in this research, available resource capacities should be known. In particular, deriving environmental capacities for a demarcated urban area can be very difficult. In addition, non-point source emission models should be applied to estimate pollution levels. Likewise, financial capacity and spending, but at the same time also revenues from, for example, road pricing, should be internalised in the modelling;

Policy measures Transport planners have several transport policy measures

available. A distinction can be made in demand-side measures (related to the affordability of travel options) as well as supply-side measures (related to the availability of travel options). To control transition paths of state variables, as guided by the transport policy objectives, these measures can be deployed (if necessary, bounded by constraints) in a prescribed sequence and timing. Transport modelling should be able to derive these paths.

Incorporating all these requirements, sustainable transport development planning is now objective-driven and resource-bounded for both current and future generations, while being infrastructure supply-limited rather than demand-driven. Hence, a proactive approach that can be characterised as one of provide - (manage -) predict (hence 'prevent'), unlike the common principle of predict - provide (- manage), remains.

7.2.3 Optimising transport policies

The transport planner has a direct role in transport policy making. Representing the decision-maker, the transport planner formulates courses of action or measures in order to solve transport problems. These measures are at best based on the knowledge of and experience in transport planning and modelling of the transport planner. Often, possible solutions are repeatedly simulated in travel demand models, evaluated and finally selected, to come up with a proposed solution (typified as a: 'What if?' strategy). Doing so, it is often almost as if the policy measures determine the policy objective, instead of the other way around.

The decision-support, provided through transport planning and modelling methods by the transport planner, forms a fundamental part in the total transport policy process, in particular the decision-making stage. Given the complex set of objectives and requirements, present-day transport planners are confronted with, as well as the complex decision-making environment the transport planner is working in, the current type of transport modelling seems less useful for this task. A decision-making tool that is capable of deriving traffic and transport measures from an explicitised sustainable transport policy objective, is therefore developed here (which may be typified as a: 'How to?' strategy).

To do so, transport planning is first characterised as an engineering design problem, where new design proposals are iteratively generated and evaluated. Each design is evaluated against some pre-defined goal and objective and if necessary modified and evaluated again until the best possible design is obtained. This is essentially optimisation, where alternatives are designed as such to make a system or design as effective or functional as possible on the basis of some pre-defined goal or objective. Therefore, transport planning has accordingly been characterised as an optimisation problem as well.

This characterisation essentially holds irrespective of the adopted decision-making approach, which may vary from a rational, normative approach to a group-

decision approach. It is only the role and use of the optimal solution that differs. In the first extreme, the solution can be regarded as the preliminary design for the transport planning, ready for further detailing and operationalisation, possibly involving other stakeholders as well. In the second extreme, optimal solutions might be derived reflecting each stakeholders objective and preferences. Those solutions will then be inputs in further negotiations, perhaps also new simulations. In addition, it is shown that in the characterisation of the design problem itself, which is the sustainable urban transport development problem, it should be possible to model most or all degrees of freedom. In other words, it is possible to add most essential elements for the problem being analysed, hence to represent the most significant aspects of the sustainable transport development design problem.

An integrated model for sustainable urban transport development thus remains that is able to assist in the complex causal reasoning from objective to a set of traffic and transport related measures. Although still being a simplification of reality, the incorporated aspects as well as the detail-level reflect the main issues involved in the strategic transport system design problem. Furthermore, all causalities modelled are straightforward, clear and explainable. Transition paths, for example, are visible, as well as the marginal costs and benefits that go with a certain sequence and timing of actions.

If solutions to the design problem appear to be infeasible or non-viable, other existing methods and models for modelling sustainable transport, which have been categorised and discussed here, perhaps used in conjunction with the dynamic optimisation model, can be consulted. For example, applying back-casting models when standard (optimised) solutions fail to reach the proposed transport policy objectives.

7.2.4 Model formulation and solution strategies

The design problem in this research has thus been characterised as a constrained dynamic optimisation problem, which accordingly determines the type and structure of the model and its equations.

In dynamic optimisation, the sequence and timing of allocating certain actions and investments, or control strategies, is determined for a certain fixed control horizon. Hence, a transport policy objective can be optimised. This is done by first predicting the exogenous and (possibly) dynamic social-economic characteristics of a population in a study area for the time horizon considered, followed by determining the associated system dynamics behaviour, given the transport network characteristics, while calculating the optimal control strategy, in order to move the system in the direction of the maximised transport policy objective. These actions are constrained by resource capacities and by natural limits of the control variables. To do these calculations, one of the most general approaches to control theory is deployed here, which is applied in continuous-time, hence allowing for different time-scales of decisions and acti-

ons to occur. A single discrete time-unit can typically be one month, a quarter or longer, because of the strategic character of the modelling.

Transport dynamics

Based on the basic building blocks of the traditional four-step transport model, several state and control variables have been formulated, which together make-up the dynamics of the model. The state variables are: travel demand in terms of link volume, infrastructure supply in terms of effective capacity or design capacity, pavement condition reflecting the level of maintenance, person throughput in the system, kilometres driven in the system and total non-point emissions in the system. The state variables are mostly formulated in terms of disequilibrium ordinary differential equations. Travel demand is considered to be elastic and a function of the trip generation characteristics of the zones in the study-area, as well as the inter-zonal composite costs of modes and routes (thus transport system performance). Furthermore, a simultaneous, doubly-constrained, trip distribution, mode-choice, route assignment model is used to distribute trips over the different options using the concept of travel utility in combination with a discrete-choice model. This utility is a function of the generalised costs that can be controlled by a transport planner using control variables. These are: expansion of capacity per unit of time (depending on the chosen strategic time horizon of decision-making), applying maintenance (in terms of pavement overlay) and raising general vehicle tax as well as parking tax in the destination zone. The first two variables change the effective capacity or design capacity, which again affect the travel time on a certain route, while the tax variables change the travel demand directly, through travel costs. The design capacity is the constructed capacity of a link, whereas the effective capacity allows for pavement deterioration to occur. Pavement deterioration is a function of the wear, which is caused by axle loading, and natural deterioration, which may be caused by climatological variables.

As these state and control variables have all been modelled in an integrated set of nonlinear ordinary differential equations, transition paths for each of the state variables are obtained for a given sequence of control variables over time. Because the user-defined transport policy objective is also known (this might be congestion minimisation, accessibility maximisation, person throughput maximisation or equity maximisation), the sequence of control variables can be chosen as such to maximise or minimise the transport policy objective function. Therefore, the control measures can be deployed, in order to steer the disequilibrium dynamics of travel demand and infrastructure supply (alike the static equilibrium in traditional transport modelling) in the direction of the maximised objective function.

Furthermore, constraints on the control variables as well as on the state variables have been implemented. The latter type of constraint is necessary for controlling the transition paths with respect to transport system performance like vehicle and person throughput as well as vehicle emissions.

It should, however, also be mentioned again that a disadvantage of using this type of modelling appeared to be that in the search process for an optimal solution only the vector with state and costate variable information remains known. This implies that extra information one might want to have, for example on mode-specific travel demand, can only be stored through the introduction of new (and computationally costly) state variables, implying new sets of ordinary differential equations.

Optimisation method

Dynamic optimisation in the framework of optimal control theory has been used in this research. Optimal control theory shows how to solve continuous-time maximisation or minimisation problems in which the objective function includes an integral and the constraints include a set of ordinary differential equations.

Several methods exist for solving the control problem and determining an optimal control strategy. An important group of methods are based on dynamic programming that can solve large kinds of optimisation problems under very general conditions, including problems with nonlinear and nonconvex objective functions as well as constraints. Yet, this method has not been used because of the combinatorial explosion due to the sequential solution of the value function. Instead, the solution to the optimal control problem has been characterised using the Pontryagin Maximum Principle. The Maximum Principle is very useful as it provides a set of necessary and sufficient conditions that are required to find the state variable trajectories, in other words transition paths for the state variables, and the optimal set of controls so as to maximise the objective function. These conditions are derived from the Lagrangean equation and also provide other important, call it economic, information on marginal valuation of the state variable at some point in time in terms of costs and benefits, through the costate variables. In addition, transversality conditions are imposed that are dependent on the final state, which, in this model, is free for most state variables, whereas fixed for the costate variables, which by definition is zero in the case of free endpoints for the accompanying state variables. Together with the initial values for the state variables these transversality conditions make-up the two-point boundary values. Hence, the numerical method finds a solution that satisfies the Maximum Principle, so indirectly solving the optimisation problem, rather than attempting to directly maximise or minimise the objective function subject to constraints.

If both the necessary and sufficient conditions are met, the global and unique optimal solution is found. Unfortunately, for the large-scale, nonlinear problem in this research, checking the sufficiency conditions is very complicated, as the concavity for the cost criterion, the adjointed equations, including costate variables, as well as the equality and inequality constraints, need to be guaranteed. Furthermore, the rank condition for the control variables should be satisfied as well. These checks have not (yet) been performed. Hence, the uniqueness of

the solution is not certain, and can only be checked on plausibility.

The Maximum Principle

A common approach has been used to numerically solve the two-point boundary value problem that remains from applying the Maximum Principle to the Lagrangean equation. The boundary value problem involves the state and costate differential equations with initial and endpoint conditions. The MATLAB procedure *BVP4c*, which is a collocation method for solving two-point boundary value problems, is deployed here. Collocation is a finite difference relaxation method that searches for an approximate polynomial solution over the entire interval. The optimal control is derived when both sets of boundary conditions are met.

Applying this method, first, an initial guess for all state and costate trajectories is made, from where the collocation method starts searching for a good solution by minimising the residual. The collocation method is rather sensitive to the initial guess, although less than the well-known shooting method for solving this kind of problems. Hence, a multi-start feature is built-in to restart a simulation with an improved initial guess obtained from the previous simulation. This is (most of the times) done simultaneously with lowering the value of the smoothing parameter that is adjoined to the cost criterion in combination with the new initial guess, in order to smoothen the bang-bang switches associated with the control variables.

The value of the control variables are derived in a separate routine, the zero-order problem, where the Karush-Kuhn-Tucker conditions are implemented. The control values are returned to the first-order problem accordingly. In its present shape, the zero-order problem is still rather simple, because in the case studies only values of independent controls were derived. Once they are made dependent on each other, for example implementing the budget constraint, this becomes much more complicated. In that case the zero-order problem becomes an optimisation by itself. This can be done applying some advanced methods of nonlinear programming. In particular, the risk of violating the rank condition becomes an issue in that case, as in the search for the optimum more constraints may become active than the actual number of control variables, hence tests and rules need to be built into the code to prevent and overrule this violation (see also the recommendations in paragraph 7.3).

Pure state inequality constraints can theoretically be adjoined to the Lagrangean in the same manner as with the mixed inequality constraints (those that contain a control variable explicitly) using Karush-Kuhn-Tucker conditions. However, they cannot be solved using a standard two-point boundary value problem solver, as the conditions only apply to the endpoint itself. Therefore, for the inequality state constraint, a penalty function is adjoined to the cost criterion that penalises exceeding of some target value, for example an environmental capacity constraint. This is done successfully for a continuous

inequality emission state constraint. Likewise, this is done for an integral endpoint equality emission constraint. An endpoint inequality emission constraint, however, was not successfully implemented.

Solving the system of nonlinear differential equations using the collocation method turns out to be possible, although the investigations also revealed that in certain cases this may be very difficult or even impossible. The main drawback appears to be the lack of robustness of the chosen method. The iterations must start close, sometimes very close, to a (local) solution in order to solve the two-point boundary value problem. In general, the transport dynamics seem to work very well, even though sufficiency conditions cannot be checked easily. Problems arise when control strategies are too much discontinuous, implying heavy switching between extreme values, which is improved by adding the smoothing terms in the cost criterion. Furthermore, nonlinear control variables, like the maintenance control variable, but also the vehicle tax and parking tax control variables in the travel demand equations are very difficult to solve, due to problems of singularity of the Jacobi matrix. Here, the initial trajectory guess appears to be very important again. A well chosen initial guess makes it possible to work with these combinations of independent controls in certain limited cases.

A sensitivity analysis for testing the stability of the solutions is not yet performed. The accuracy of the solution, however, can easily be tested using the initial values from the optimal solution of the boundary value problem in a standard initial value problem for systems of nonlinear differential equations (as is done for the cases where the maximum residual didn't meet the required accuracy).

To improve user-friendliness, the MATLAB code has been implemented as such that all partial derivatives for the state and costate equations are automatically derived from incidence tables. These tables can be easily constructed for each transport network and should contain information on origin-destination - route - mode - link incidences, as well as on road link characteristics and zonal trip generation information.

Application

For some synthetic case studies (a corridor model as well as two network models), different combinations of travel demand equations and infrastructure supply equations, alongside with several control variables, have been tested and demonstrated. Hence, it is possible to study the actual dynamic behaviour of the model in certain cases, to derive implications and possibilities for sustainable transport research as well as to conclude on drawbacks associated with the chosen and implemented optimisation method. As there is a large number of model variations possible, some arbitrary, but increasingly complex cases in terms of traffic dynamics, number of controls, constraints, type of objective function as well as network sizes have been implemented.

The case studies show that the dynamic model, in its present form, can be applied directly to some restricted cases with strategic networks of limited numbers of (aggregated) zones and (aggregated) links. The dynamic model reveals control paths for achieving a sustainable and developed transport system, while also providing information on the values for the state variables (including those for the system induced effects) and their associated costate variables over time. The control paths appear to be of a bang-bang structure most of the times, first starting at an upper-bound for some time, then going back to the lower bound once and for all. Hence, the control sequence and timing is revealed, which is different depending on the chosen transport policy objective.

For each policy objective, a different control scheme can be observed. The congestion minimisation objective can be seen to specifically allocate resources to individual links in order to steer the link-based level-of-service criterion, irrespective of the ‘importance’ of the origin-destination pair that is being served by these links. The accessibility maximisation objective focusses more on specific corridors (in particular when bounded by an environmental capacity state constraint) that serve most origin-destination attractions. The person throughput objective, in addition, focusses more on the trip makers themselves and allocates pro-actively to links that serve higher-occupancy vehicles. Lastly, the equity objective shows to allocate resources to links that are being used by certain population segments more than by other segments in order to keep their individual differences in experienced composite costs minimal. No final judgement on the ‘sustainability’ of the different objectives has been made yet.

From the case studies, however, it can also be concluded that it will be rather difficult to fully implement the dynamic transport optimisation model in the way it is set-up now. In particular, there where several interdependent control variables have been introduced, for example when implementing the budget constraint, computational problems arise. Summarising, the following main problems have been encountered in the case studies:

1. The computation of optimal bang-bang control due to linearity of the control variable appears to be difficult. Hence, perturbed energy terms, or smoothing terms have successfully been added;
2. For the nonlinear control variables, the Jacobi matrix is singular on a large domain. Well-chosen initial guesses make it possible to derive some optimal controls though. For some of the control variables, however, difficult nonlinear zero-order problems remain present. Some more advanced solution techniques should be incorporated to tackle this problem;
3. Checking the sufficiency conditions associated with the Pontryagin Maximum Principle is not easy for this problem. Hence, the optimal solution is not necessarily, the unique and global one;
4. Introducing constraints alike the budget constraint, the zero-order problem becomes more difficult because an extra constraint (without incre-

asing the number of control variables) is added to the problem, which increases the risk of rank condition violation.

It appears that numerical methods for solving optimal control problems have not yet reached the stage that for example methods for solving differential equations have nowadays. Solving an optimal control problem, requires significant user involvement in the solution process. In the recommendations for further research some discussion on areas, where the dynamic model and the solution methods have to be improved and suggestions on how to do so, is given.

Because of this apparent lack of robustness of the chosen solution method, the model has not been implemented using a full set of real world data yet. Therefore, no real policy implications from the example case studies can be drawn. However, from the initial development of methods to solve the complicated model, i.e. the case study results, it may not be long before a full model (perhaps after some alterations in the model formulation or solution method, as is discussed in the recommendations) can be implemented, estimated and applied in strategic transport planning. In particular if the dynamic model is going to be used in a hierarchical modelling, which is in a combination with a lower-bound traditional transport model for operationalising the strategic set of transport measures, this will greatly improve transport planning and modelling and bring a sustainable and developed transport system closer to reality.

7.2.5 Sustainable urban transport development

From the present conceptual development, the insertion within the transport policy process as well as the model building in this thesis, it can be concluded that the main aim of this research, which is to demonstrate the implications of a transport planning paradigm shift and to develop a corresponding analytical framework involving change from a given state of the transport system to a system compatible with sustainable development, has been achieved. The paradigm of sustainable urban transport development proved to require that sustainable transport policy objectives become determinative in transport planning, more than they do now. An explicit thinking from transport policy objectives towards sets of transport measures, including a set of rules on the sequence and timing of these optimal measures, has thus been advocated and (conditionally) made possible.

The dynamic modelling and applications presented here are believed to provide a good step in developing a method and tool that can actually be applied in practice, even though some computational problems with the modelling have been encountered that need to be resolved at some point. It should also be noted that the translation, call it operationalisation, of the conceptual framework for sustainable urban transport development into a dynamic transport model, has only started. Different requirements, in particular those related to non-motorised transport and public transport options, especially in relation to the transport realities in developing countries, should be internalised in the model

as well. Eventually, also land-use interaction with transport, in particular due to the relatively long time-scales of strategic modelling, should be internalised, in order to fully operationalise the concept.

The three following main and summarising conclusions, directly answering the research questions, may thus be drawn:

What are the implications of the notion and definition of sustainable development for urban transport planning? The notion and definition of sustainable development are integrated into urban transport planning by advocating a comprehensive decision-making that anticipates and manages scarce resource use, while developing the transport system in terms of quality of access and/or person throughput, hence stimulating social-economic development, based on an explicit and integrated transport policy objective;

How should sustainable development be modelled and incorporated in urban transport planning practice? Internalising the notion and definition of sustainable development into transport modelling is only possible if it is seen and characterised as a dynamic and constrained optimisation problem, while incorporating a quantified transport policy objective as well as resource bounds. Doing so, the requirements that follow from a conceptualisation of sustainable urban transport development are directly integrated into a supply-limited, in other words provide-manage-predict, transport modelling that is readily applicable in urban transport planning practice;

What typical consequences can be expected and implications be drawn for urban transport planning in the long-term due to internalisation of the concept of sustainable development? Applying the concept of sustainable urban transport development, through the use of a dynamic optimisation model, shows where, when and by how much transport measures should be applied in order to steer the system in the direction of the maximised sustainable transport policy objective. A more effective and efficient (scarce) resource allocation can be observed. Because of the dynamics in the model there is continuous feedback between the different sub models of the transport model, hence also incorporating generated demand. Implications of measures in terms of transport system performance, but also the number of generated trips in time and total emissions, are thus known. In addition, ‘marginal costs’ associated with erecting boundaries to some control and state variables are explicitly known to the transport policy maker.

Besides, it should be noted that the presented framework as well as the dynamic modelling can both serve as a basis for studying other, but similar, type of transport problems, which are not necessarily related to issues of sustainability or transport planning. For example, an application involving traffic control decisions for short-term traffic management may also be based on the same principles.

Some ideas on the necessary further development of the concept and its operationalisation are discussed next.

7.3 Further research

In this paragraph, some perspectives on possible future research are provided. First, some recommendations are given with respect to the characterisation and conceptualisation of the sustainable urban transport development problem itself, followed by recommendations that deal with the formulation of the dynamic model. In addition, some recommendations with respect to the chosen solution method are discussed, making a distinction between issues related to model accuracy and issues related to alternative formulations. Finally, recommendations on possible future applications and case studies are given, in particular with respect to the issue of data collection.

7.3.1 Characterisation of the problem

In this research, the paradigm of sustainable urban transport planning has been formulated mainly in strategic terms. The concept should be extended for use on a tactical as well as an *operational level* of sustainable transport decision-making. This can be done by more explicitly incorporating essential issues like parking solutions, local air pollution, different types of vehicles, traffic calming etceteras. Many of these solutions can for example be found in several articles by Todd Litman (see bibliography).

Further research is now being conducted on defining *environmental capacities* for a demarcated urban area, amongst others incorporating the ecological footprint concept with these environmental capacities as well, extending the work of Van Nes (2004). Perhaps also adopting some of the knowledge in the earlier mentioned papers by Huapu and Peng (2001) and Shrestha et al. (2005). Well defined environmental capacities can then be used as pure state emission constraints in the model.

In addition, more attention should be given to estimating *travel demand elasticity*. Although it has been shown in many researches that a considerable amount of generated demand exists, a generally applicable framework for incorporating this in transport methods and models is non-existing, as found and discussed in Bierman (2004).

The concept of *equity* is of great importance to the concept of sustainable urban transport development. A quantification of equity, however, appears to be very difficult. More effort should be put in defining equity and translating this concept in modelling terms.

Land-use transport interaction is another important aspect to the concept of sustainable urban transport development. Hence, this interaction should also be integrated in the further development of the concept, preferably adopting

some of the findings in the development of an appraisal model for urban land-use and transport strategies, perhaps also the computational method as discussed and applied in Pfaffenbichler (2003). Furthermore, some of the wisdom on land-use transport dynamics mentioned in Tillema (2004) might be adopted.

7.3.2 Model formulation

In this research continuous-time optimal control formulations have been used, which is reflected in the type of equations being used. However, some further efforts are necessary to improve some of the model equations, while incorporating the suggested extensions in the characterisation above.

The *simultaneous distribution, mode-choice, assignment model*, used here, is based on applying utility functions in a logit-model formulation. However, this particular sub model has also been criticised in literature as it ignores correlations amongst alternative choices. Therefore, additional research should be conducted in improving this part of the model, for example investigating a continuous nested-logit formulation that partitions the choice set in several nests or probit models that explicitly capture correlation amongst alternatives. Furthermore, the *travel demand elasticity* formulation, currently calculated as a revealed demand ratio per corridor, might be improved, for example by adapting insights presented in Bell and Iida (1999).

No equations have yet been formulated that explicitly deal with *non-motorised* and/or *public transport* modes. Instead, equations have been formulated in terms of general effective capacities (so not necessarily excluding these mode categories!). Within the concept of sustainable urban transport development both mode-categories ought to have a more prominent role. This can be done by, for example, adopting some of the wisdom on modelling public transport priority lanes in continuous-time, as described in Donaghy and Schintler (1994).

Pavement deterioration characteristics, travel behaviour as well as effective road capacities in most *developing cities* will significantly differ from developed cities. Operationalising the concept of sustainable urban transport development can only be done with a thorough understanding of these particular circumstances. Hence, further research should be dedicated on exploring these relations. Good descriptions on transport realities in these cities are given in Vasconcellos (2001) and World Bank (2002).

Car ownership models, population segmentation models, advanced traffic assignment algorithms, etceteras, have been integrated in standard transport models nowadays. Knowledge of these sub models should be used to upgrade some of the continuous equations in the model presented here. Again, use can be made of some of the sub models applied in Pfaffenbichler (2003).

In terms of externalities, *energy consumption models*, which represent the energy consumption rate, hence relating speed and fuel consumption, should be

added to the model, for example following Zietsman (2000). In addition, incorporating a *noise pollution model* (in particular for use at a more operational level), as expressed in chapter 1 might be considered. Furthermore, it is advised to further evaluate the applied *non-point source emission model* and see whether and to what extent it is justified to use one dominant pollutant instead of using separate models for each pollutant.

The implementation of more and more complex (nonlinear) mixed inequality constraints, in particular constraints like the *budget constraint*, should be studied further, as these are also crucial to the concept of sustainable urban transport development. Rank condition violation becomes an issue here, hence tests need to be built-in and strategies for overruling the violation need to be provided. Working with (sets of) nonlinear controls, as is also the case here with respect to the vehicle tax as well as parking tax control variables, requires, in addition, that more attention needs to be paid to a well-chosen *initial guess* for the costate variables. Therefore, it should be studied how this initial trajectory guess can be made easier, or whether for example linearising functions might be a valid and useful option.

7.3.3 Optimisation

Accuracy

Further attempts should be made to test the *sufficiency conditions* that come with Pontryagin's Maximum Principle.

In addition, *sensitivity analysis* should be conducted in particular with respect to the lagged-adjustment variables as well as to some of the scale parameters in the logit model and composite cost model. This because it will be difficult - but not impossible - to estimate these parameters.

Furthermore, some of the test-results can be double-checked by taking the optimal initial values for the costate variables (from the solution of the boundary value problem) in combination with the initial values for the state variables and use them as initial values in an *initial value problem* (as is done for some cases already). This will give more insight in the accuracy of the chosen solution method.

Alternative methods

It is recommended to further investigate the possibilities for solving the optimal control model by applying different *computational techniques* for the indirect solution method, like shooting methods as discussed in paragraph 6.2. Furthermore, it should be tested whether solving the optimal control by means of dynamic programming could improve the model accuracy and outcomes. In addition, and in combination with dynamic programming, solving a discrete-time optimal control might be considered.

Another model formulation in terms of a *closed loop control*, instead of the open loop control applied here, can be considered. In the case of closed loop control, the control variables are functions of the state variables directly instead of time, hence obtaining a so-called optimal feedback law, with closed-loop trajectories for the state variables satisfying some optimality conditions. Solution techniques in this case are based on dynamic programming as well. Furthermore, the use of direct computational methods, where the optimal control problem is transformed into a finite dimensional optimisation problem, might be investigated, as direct methods feature a larger area of convergence. However, their overall convergence rates are rather slow, as is described in Schwartz (1996).

On the other hand, minor and major changes in the model equations (perhaps implying changes to the conceptual framework itself!) can be considered. Barten (2004), for example, provides some *alternative formulations* to the optimisation model discussed here. A welfare function is for example introduced as an objective function, alike the formulation used in a publication by the European Commission (1998). Alternatively, in view of the problems mentioned before, an obvious change would be to make the model static, hence getting the possibility to solve an alternative model, which is not necessarily complying with the sustainable urban transport development paradigm, by techniques of combinatorial optimisation, nonlinear programming, multi-objective programming, bi-level programming, variational inequalities etceteras. Most of these methods have been successfully applied to network optimisation problems. Particularly, because they hardly suffer scale problems due to network size.

7.3.4 Future application

Apart from the fact that a full-scale real-life application to a strategic transport network would show full potential of the concept as well as the modelling, it is first recommended to study the problems that arise with respect to *data collection* issues. In particular, those related to the dynamics of the model.

To fully estimate the model, *longitudinal transport data* needs to be collected. As transport systems cannot be confined to laboratory environments, collecting data over time can be very difficult. That is, panels of respondents need to be followed and detailed observations of transport infrastructure changes over time need to be made. Hence, other experiences with longitudinal data collection, like those summarised in Goodwin et al. (1990), need to be studied before a real case study with accompanying data collection can be performed. Observing longitudinal travel patterns as well as travel mechanisms related to non-motorised transport and public transport, particularly in *developing cities*, might pose some interesting research challenges.

The same applies to collecting data as well as to parameter estimation with respect to travel demand elasticity. In the course of this study, estimating such parameters by means of cross-sectional data (comparing social-economic equivalent, but in accessibility terms non-equivalent groups, by means of a hou-

sehold survey) for Metro Manila, The Philippines, proved to be very difficult, as is described in Bierman (2004). Hence, also with respect to estimating travel demand elasticity, longitudinal data appears to be needed.

Furthermore, the typical time-scales of adaption for the various state variables, reflected through the *lagged-adjustment parameters*, should be obtained using longitudinal data.

Although the model has been prepared with interpolation algorithms for *exogenous zonal social-economic variables*, these have not been used to their full-extent in the present case studies. Knowledge on dynamics of exogenous variables should therefore be implemented in more detail than is done now.

In a strategic transport model application, usually a relatively small number of zones and links is used. These zones and links might be obtained using *zonal and network aggregation techniques*. The aggregation of transport networks in particular, should be automated because in 'standard' size transport networks, the number of links is in hundreds or more. Research is underway for aggregating such networks, as is reported in Van Gent (2005).

Furthermore, any future application should be done in conjunction with a *baseline (do-nothing) scenario*, as well as an operationalisation in a hierarchical model structure.

Chapter 8

Epilogue

In this thesis no distinction has been made between urban transport in either developed or developing countries, as the urban transport problem is regarded to be rather universal. The specific manifestation of the urban transport problems may differ between these cities, but the underlying mechanisms, however, which have been studied here, are basically the same. Yet, the magnitude of these problems, particularly in terms of levels of congestion and pollution as well as problems of accessibility and traveller discomfort, is frequently much worse in developing cities than in cities in the developed world. Hence, in this short epilogue some more perspective will be given to the problems in developing countries and the possible role of the paradigm and subsequent modelling for sustainable urban transport development, which have been developed here.

Urban areas in the developing world are continuously and rapidly expanding in size and population density. With this growth, transport problems, in particular those related to congestion and (local) environmental pollution, for example because of rapidly rising levels of car ownership, have often become unmanageable. At the same time, these cities can hardly survive or develop in social-economic terms, if the transport system is maintaining an equilibrium demand - supply level, whereby the city gets more and more congested and polluted, so inhabitants and trip makers experience increasing levels of discomfort. Lacking transport facilities and bad transport planning might be few of the reasons, while the transport system ought to provide the essential means to accessibility of housing, jobs, education, health care etceteras. This because the availability of adequate road infrastructure is a (necessary, but not sufficient) prerequisite for social-economic development.

In view of the issues of sustainability and sustainable development with respect to transport as discussed in this thesis, a better understanding of transport realities and travel behaviour in these cities, the nature and magnitude of transport problems of the urban poor, and the decision-making framework and practice in their societies, is essential for proposing measures and solutions in transport

plans aiming at an improved transport system. Major transport differences exist in terms of travel behaviour patterns (e.g. due to population density, unemployment rates of city dwellers, spatial setting of planned and unplanned (squatter) areas, informal employment and other employment characteristics), in terms of the wide variety of transport means available (e.g. walking, bicycles, tricycles, rickshaws, jeeps, shared-taxis, private cars, taxis, buses, handcarts, trucks, pushcarts), and their accompanying ownership structure, but also in terms of the infrastructure available (ranging from dirt tracks to six-lane expressways with (very) mixed traffic), not even mentioning common problems of traffic behaviour and enforcement as well as equity in the distribution of transport opportunities available to the people.

In addition, some of the assumptions done in common transport planning models as well as in the dynamic model that is proposed in this research, notably the present inadequacy of both type of models to adequately deal with non-motorised travel, don't comply with these specific conditions. At the same time, the current ability of the dynamic model to explicitly address the transport policy question, furthermore to deal with dynamic changes in travel patterns and behaviour, moreover to include induced demand, while realising that this induced travel offsets emission savings and accessibility gains, all three very important and relevant in developing cities, reveal a good prospective use of the framework and model. However, this can only be done if a better understanding of the context in which the methods and models have to operate as well as of the function the model should fulfill in the transport planning and policy process, is obtained. Moreover, as data collection and techniques of data collection are continuously improving, also in developing countries, there is enough ground for seeing the importance of applying the proposed method and model of transport planning in the future.

Furthermore, it should be noted that road infrastructure expansion as well as management of infrastructure targeted at reducing congestion, improving accessibility and person throughput, while keeping track of fuel consumption and pollutant emissions capacities, which is the main focus in this thesis, is just one of the significant dimensions in transport sustainability in developing cities. However, these may appear to be essential instruments for making urban transport more efficient and sustainable, depending on the situation. In those places where an adequate road network is lacking, this might even be the only worthwhile option.

Doing so, it is very likely that the equilibrium demand - supply level can be steered as such as to meet an explicit transport policy objective as well as pollution targets, while at the same time allowing more people to make their trips in relative comfort using motorised, non-motorised as well as public transport. This, because a well prescribed sequence and timing of allocating engineering and pricing actions has been performed. Combined with an overall and integrated urban vision for sustainable development, as is already rightly done in

some cities, for example the city of Bogota, Colombia, a rather different set of investment priorities, better reflecting the current and future urban travel related needs, is expected.

-/-

Appendix A

Adverse effects from vehicle emissions

Increases in road traffic have produced unsustainable levels of congestion and pollution, although some emission types that affect air quality, like nitrogen oxides (NO_x) and non-methane volatile organic compounds (VOCs), are now reducing at least in Europe, as could be seen in figure 1.2. The effects of traffic, however, can still be felt at a local level through poor air quality, noise, severance and at a global level through climate change. Vehicles emit significant amounts of several pollutants with varying effects. This appendix gives an overview of the major air pollutants and discusses the emission of greenhouse gasses and its effects separately.

A.1 Air pollutants

Major type of air pollutants include particulate matter, lead, carbon monoxide, nitrogen oxides, volatile organic compounds and hydrocarbons as well as photochemical oxidants such as ozone. Table A.1, taken from Schwela (2002), exhibits selected pollutants and their major sources and effects.

In developing cities, the most critical air pollutants are particulate matter and lead (if not yet phased out of petrol). Their concentrations of air pollution often reach levels of concern for public health.

The term volatile organic compounds, as mentioned before and applied in chapter 6, refers to organic compounds that readily evaporate. VOCs, include hydrocarbons (HC), and organic compounds containing chlorine, sulphur or nitrogen. VOCs, in combination with NO_x , are responsible for ground level ozone and smog.

Table A.1: Sources of air pollution, effects, and WHO guidelines for selected pollutants (Schwela, 2002)

Pollutant	Major sources	Effects	Health guidelines
Carbon monoxide (CO)	Motor-vehicle exhaust; some industrial processes	Poisonous to humans when inhaled. CO reduces the oxygen carrying capacity of blood and places additional strain on the heart and lungs	10 mg m ⁻³ (10ppm) over 8h; 30 mg m ⁻³ over 1h (30,000 µg m ⁻³)
Sulphur dioxide (SO ₂)	Minor contribution from mobile sources, Heat and power generation facilities that use oil or coal containing sulphur; sulphuric acid plants	A human irritant, SO ₂ undertakes atmospheric reactions to contribute to acid rain	125 µg m ⁻³ over 24h; 500 µg m ⁻³ over 10 min
Particulate matter PM ₁₀	Soil, oceanic spray, bush fires, domestic burning, motor vehicles, industrial processes and organic dusts from plant material	Contribute to haze, increases cancer risk, mortality effects, aggravates respiratory illnesses	Available information does not indicate concentrations below which no effect would be expected. For this reason no guideline value for short term average concentration is recommended
Lead (Pb)	Added to some fuels, Pb is emitted from motor-vehicle exhausts; lead smelters; battery plants	Affects intellectual development in children; many other adverse effects	0.5 µg m ⁻³ over a year
Nitrogen oxides (NO, NO ₂)	A side effect of high combustion temperatures causing bonding of nitrogen and oxygen in motor vehicle exhaust; heat and power generation; nitric acid; explosives; fertilizer plants	Irritant, precursor to photochemical smog formation	200 µg m ⁻³ over 1h for NO ₂
Photochemical oxidants (primarily ozone (O ₃); also peroxyacetyl nitrate (PAN) and aldehydes	Formed in the atmosphere by reaction of nitrogen oxides, hydrocarbons, and sunlight	An irritant, photochemical oxidants contribute to haze, material damage, and, aggravates respiratory illnesses	120 µg m ⁻³ over 8h

A.2 Climate change

Greenhouse warming¹ occurs when certain gases allow sunlight to penetrate the earth but partially trap the planets radiated infrared heat in the atmosphere. Some such warming is natural and necessary. If there were no water vapour, carbon dioxide, methane, and other infrared absorbing (greenhouse) gases in the atmosphere trapping the earths radiant heat, our planet would be about 60 degrees F (33 degrees C) colder, and life in its current way would not be possible. Naturally occurring greenhouse gases include water vapour, carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O), and ozone. Several classes of halogenated substances that contain fluorine, chlorine, or bromine are also greenhouse gases, but they are for the most part, solely a product of industrial activities. Chlorofluorocarbons (CFCs) and hydrochlorofluorocarbons (HCFCs) are halocarbons that contain chlorine; while halocarbons that contain bromine are referred to as halons. Other fluorine containing halogenated substances include hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), and sulfur hexafluoride (SF₆). There are also several gases that, although they do not have a direct global warming effect, do influence the formation and destruction of ozone, which does have such a terrestrial radiation absorbing effect. These gases include carbon monoxide (CO), oxides of nitrogen (NO_x), and nonmethane volatile organic compounds. Aerosols, extremely small particles or liquid droplets often produced by emissions of sulfur dioxide (SO₂), can also affect the absorptive characteristics of the atmosphere. Although CO₂, CH₄, and N₂O occur naturally in the atmosphere, the atmospheric concentration of each has risen, largely as a result of human activities. Since 1800, atmospheric concentrations of these greenhouse gases have increased by 30%, 145%, and 15%, respectively. This build up has altered the composition of the Earths atmosphere, and may affect the global climate system. Beginning in the 1950s, the use of CFCs and other ozone-depleting substances (ODSs) increased by nearly 10% a year, until the mid-1980s when international concern about ozone depletion led to the signing of the Montreal Protocol. Since then, the consumption of ODSs has rapidly declined as they are phased-out. In contrast, use of ODS substitutes such as HFCs, PFCs, and SF₆ has grown significantly; all of which have strong greenhouse-forcing effects. In late November 1995, the Intergovernmental Panel on Climate Change (IPCC) Working Group 1 concluded, ‘the balance of evidence suggests that there is a discernible human influence on global climate.’ The transportation sector is responsible for approximately 17% of global carbon dioxide emissions and these emissions are increasing in virtually every part of the world. The potential global warming benefits of diesel vehicles, due to their substantial fuel economy benefits relative to gasoline-fuelled vehicles, have been undercut by recent studies. These indicate that diesel particles may, by reducing cloud cover and rainfall, more than offset any CO₂ advantage. As noted by NASAs James Hansen, ‘Black carbon reduces aerosol albedo, causes a semi-direct reduction of cloud cover, and reduces cloud particle albedo.’ Eutrophication Nitrogen oxides also result

¹This paragraph is taken from Asian Development Bank (2003).

in nitrogen deposition into sensitive nitrogen-saturated coastal estuaries and ecosystems, causing increased growth of algae and other plants. Long-term monitoring in the United States, Europe, and other developed regions of the world shows a substantial rise of nitrogen levels in surface waters, which are highly correlated with human-generated inputs of nitrogen to their watersheds. Fertilisers and atmospheric deposition dominate these nitrogen inputs. Human activity can increase the flow of nutrients into those waters and result in excess algae and plant growth. This increased growth can cause numerous adverse ecological effects and economic impacts, including nuisance algal blooms, die-back of underwater plants due to reduced light penetration, and toxic plankton blooms. Algal and plankton blooms can reduce the level of dissolved oxygen, which adversely affect fish and shellfish populations. This problem is of particular concern in coastal areas with poor or stratified circulation patterns. In such areas, the overproduced algae tends to sink to the bottom and decay, use all or most of the available oxygen and thereby reduce or eliminate populations of bottom-feeder fish and shellfish, distort the normal population balance between different aquatic organisms, and in extreme cases, cause dramatic fish kills. Collectively, these effects are referred to as eutrophication.

Appendix B

Models estimation

B.1 General least squares criterion

In least squares estimation¹, the unknown values of the vector of parameters: $\vec{\zeta} = \zeta_0, \zeta_1, \dots, \zeta_n$, in the regression function, with the vector of independent explaining variables: $\vec{x} = x_0, x_1, \dots, x_n$, $f(\vec{x}, \vec{\zeta})$, are estimated by finding numerical values for the parameters that minimise the sum of the squared deviations between the observed responses y_i and the functional portion $f(\vec{x}; \vec{\zeta})$ of the model. Mathematically, the least (sum of) squares criterion that is minimised to obtain the parameter estimates is

$$S = \sum_{i=1}^n [y_i - f(\vec{x}; \vec{\zeta})]^2. \quad (\text{B.1})$$

The intercept ζ_0 and the other elements of $\vec{\zeta}$ are treated as the variables in the optimisation and the independent variable values, \vec{x} , are treated as coefficients. To emphasise the fact that the estimates of the parameter values are not the same as the true values of the parameters, the estimates are denoted by $\vec{\zeta}$. For linear models, the least squares minimisation is usually done analytically using basic calculus. For nonlinear models, on the other hand, the minimisation must almost always be done using iterative numerical algorithms. A distinction is often made between linear regression and multiple regression models. The latter has more explanatory, independent, variables and consequently more regressors in $\vec{\zeta}$. The solution equations are similar, although more complex.

The linear regression is illustrated here for a straight-line model. Consider:

$$y = \zeta_0 + \zeta_1 x + \epsilon, \quad (\text{B.2})$$

¹The first part of this appendix, introducing least squares estimates up to equation (B.5), is based on NIST-SEMATECH (n.d.).

where y is the dependent variable, x the independent variable, ζ the optimising variables and ϵ a disturbance term to describe the deviation of y from its expected value. For this model the least squares estimates of the parameters can be computed by minimising:

$$S = \sum_{i=1}^n \left[y_i - (\hat{\zeta}_0 + \hat{\zeta}_1 x_i) \right]^2. \quad (\text{B.3})$$

This can be done by:

1. taking partial derivatives of S with respect to $\hat{\zeta}_0$ and $\hat{\zeta}_1$;
2. setting each partial derivative equal to zero;
3. solving the resulting system of two equations with two unknowns,

which yields the following estimators for the parameters:

$$\hat{\zeta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (\text{B.4})$$

and:

$$\hat{\zeta}_0 = \bar{y} - \hat{\zeta}_1 \bar{x}. \quad (\text{B.5})$$

These formulas are instructive because they show that the parameter estimators are functions of both the independent and dependent variables and that the estimators are not independent of each other unless: $\bar{x} = 0$. This is clear because the formula for the estimator of the intercept depends directly on the value of the estimator of the slope, except when the second term in the formula for $\hat{\zeta}_0$ drops out due to multiplication by zero. This means that if the estimate of the slope deviates a lot from the true slope, then the estimate of the intercept will tend to deviate a lot from its true value too. This lack of independence of the parameter estimators, or more specifically the correlation of the parameter estimators, becomes important when computing the uncertainties of predicted values from the model.

The coefficient of determination, indicating the accuracy of the regression estimate, is often depicted as the ratio between the explained variation ($\hat{y}_i - \bar{y}$), where: $\hat{y}_i = \hat{\zeta}_0 + \hat{\zeta}_1 x_i$, is the estimated value of the dependent variable y_i and unexplained variation ($y_i - \hat{y}_i$), or:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (\text{B.6})$$

R^2 has limited values of 1 in the case of perfect explanation and 0 if no explanation at all. Intermediate values may be interpreted as the percentage of the total variation explained by the regression.

In case of multiple regression models, multiple k variable values in \vec{x} , hence k regressors in $\vec{\zeta}$, exist. As adding each regressor automatically increases the R^2 value, the *corrected* or *adjusted* R^2 is defined as:

$$\tilde{R}^2 = \frac{(R^2 - k)}{(n - 1)} \frac{(n - 1)}{(n - k - 1)}, \quad (\text{B.7})$$

where n is the sample size and k the number of variables or regressors. Hence, for multiple regression it is better to use the adjusted coefficient of determination.

Furthermore, one should be cautious for multicollinearity, which occurs when there is a linear relation between two explanatory variables, like income and occupation for example. The regressors are not independent anymore.

For a more detailed explanation the reader is, for example, referred to Ortúzar and Willumsen (2001).

B.2 Maximum likelihood estimation

If one observes a random sample: x_1, x_2, \dots, x_n , of values drawn independently from a distribution, say $f(\vec{x}|\theta)$ governed by an unknown parameter θ , the probability of obtaining the data may be expressed given a value of θ . This expression is called the *likelihood* of the sample (or likelihood on the data). It may be thought of as the chance of obtaining the sample data that is actually obtained given θ . Because θ is usually unknown, it must be estimated from the data. As an estimate of θ , the value $\hat{\theta}$ is selected as such that when evaluated at $\hat{\theta}$, the expression for the likelihood of the sample reaches a maximum. This process of finding estimated values of unknown parameters is called maximum likelihood estimation (MLE)². Estimates obtained in this manner are known as maximum likelihood estimates (MLEs).

The principle aim of MLE is to find parameter values that maximise the sample likelihood, L , which may be thought of as the formula for the joint probability distribution of the sample, which in the case of the multinomial logit model applied in this research is a logistic distribution. The *likelihood function* yields a value that is proportional to the joint probability (or likelihood) of obtaining the particular data that are actually observed. Assuming independent observations, the individual components of the likelihood function can be multiplied using the general rule for joint probabilities of independent events.

If \vec{x} denotes a random sample of n independent observations from a population with a density function, $f(\vec{x}|\theta)$, where θ is an unknown parameter describing

²Discussions on Maximum Likelihood Estimation, like this one, can be found in any textbook on econometrics or advanced statistics, but also in books on discrete-choice models like Ben-Akiva and Lerman (1985).

some aspect of the population distribution of \vec{x} . The joint density function of the sample is the product of the individual density functions, given by the expression:

$$L = \prod_{i=1}^n f(x_i|\theta). \quad (\text{B.8})$$

The goal of MLE is to find the set of values of the unknown parameters that make L as large as possible. Instead of maximising the likelihood function, it is often more convenient to maximise the *log-likelihood function*. Because the logarithm is a strictly monotone transformation, the set of values that maximise L will also maximise $\log L$, in this case:

$$\log L = \sum_{i=1}^n \log f(x_i|\theta). \quad (\text{B.9})$$

For this example, which involves only a single parameter, the maximum is attained when the rate of change of $\log L$ with respect to θ equals zero. This condition is referred to as a first order condition. Mathematically, this condition is expressed by equating the first partial derivative of $\log L$ with respect to θ , to zero and solving for θ accordingly:

$$\frac{\partial \log L}{\partial \theta} = 0. \quad (\text{B.10})$$

The solution of this equation yields the MLE for $\hat{\theta}$. To ensure that $\log L$ is maximised when solving for θ , it must be the case that the slope of $\log L$ is decreasing near the MLE. This condition is called the second order condition, given by the expression for the second partial derivative of $\log L$ with respect to θ :

$$\frac{\partial^2 \log L}{\partial \theta^2} < 0. \quad (\text{B.11})$$

This solution can be obtained by a numerical optimisation algorithm, for example applying an iterative procedure as the Newton-Raphson method, which is discussed for MLEs in amongst others Ben-Akiva and Lerman (1985).

B.3 Estimating parameters in nonlinear differential equation systems

Precise estimation of model parameters is essential for having proper dynamic models. Most existing methods for parameter estimation of nonlinear systems

rely on simplifying assumptions to obtain a tractable but approximate solution for the system, with and without constraints. They often rely on assumptions about the nature of the model or the probability distributions of the underlying variables to obtain a tractable optimisation problem. The general identification problem is shortly introduced here.

If the dynamics of the model can be described by a system of ordinary differential equations, as is the case in the model in this research, then:

$$\frac{dx}{dt} = f(x, p, t), \tag{B.12}$$

where the right-hand side f depends on an unknown vector of parameters p . It is assumed that there is a possibility to do field observations measuring a ‘signal’ $\eta(t_j)$ of state variables at discrete points in time, t_j , with $j = 1, \dots, k$. The ‘output’ signal of the dynamic system is $q(x(t_j), p, t_j)$. If the modelling error is $\epsilon(t_j)$, then:

$$\eta(t_j) = q(x(t_j), p, t_j) + \epsilon(t_j). \tag{B.13}$$

It is quite common, in order to determine the unknown parameters to solve the optimisation problem by minimising a special functional under constraints that describe the specifics of the model. Any norm of the measurement error may be used as the functional in the optimisation problem. The type of the norm may be determined by the statistical distribution of the measurement error. If the errors are independent, normally distributed with zero mean and known variances σ_j^2 , minimising a general weighted least squares function, equals:

$$\min_p \sum_j \frac{(\eta(t_j) - q(t_j, x(t_j), p))^2}{\sigma_j^2}, \tag{B.14}$$

which yields a MLE (see paragraph B.2), provided that the measurement errors are independent and normally distributed with constant standard deviation σ_j .

To estimate the parameters in the dynamic model of this research advanced techniques of estimation may have to be applied (like Kalman filters, if random variables are concerned that is when stochastic ordinary differential equations are used or (constrained) maximum likelihood techniques or (constrained) linear or nonlinear least squares techniques for deterministic, but unknown, variables). An extensive description of these methods can be found in Kay (1993). In MATLAB several procedures for applying these techniques are available. In addition, Wymer (2001) provides in his software WYSEA a set of programs for the estimation, dynamic analysis, forecasting and simulation of difference or differential equation systems, where particular attention has been given to numerical precision within the whole system; thus some algorithms are used for precision rather than for speed, which is also reported in Wymer (1992).

Furthermore, the data collection for $\eta(t_j)$ itself, in particular the collection of longitudinal time series for transport, that is to follow the mobility and accessibility of a group of individuals over a period of time, including analysing the policy variants, hence their effects, over time, poses some challenges. Longitudinal data in transport research is discussed in amongst others Goodwin et al. (1990).

Appendix C

Further sufficiency conditions

C.1 Concave and convex functions

A function f is concave¹ if:

$$f(ty + (1-t)z) \geq tf(y) + (1-t)f(z), \quad (\text{C.1})$$

for each pair of points (y, z) and for all $t \in [0, 1]$. Similarly, a function f is convex if $-f$ is concave.

If f is a concave function of one variable, say x , and is differentiable, then:

$$\frac{\partial^2 f(x)}{\partial x^2} \leq 0, \quad (\text{C.2})$$

and

$$f(z) \leq f(y) + \frac{\partial f(y)}{\partial x}(z - y), \quad (\text{C.3})$$

for each pair of points (y, z) .

This is illustrated in figure C.1 with a geometrical interpretation for a function of one variable. Namely, $f(x)$, defined on interval $[a, b]$, is concave, if for each pair of points on the graph of $f(x)$, the line segment joining these two points lies entirely below or on the graph of $f(x)$.

¹Discussions on convexity and concavity like this one can be found in any text-book on calculus or dynamic optimisation.

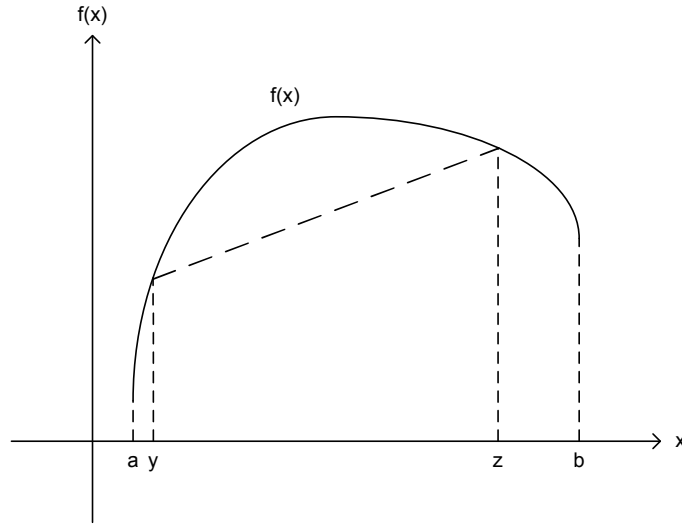


Figure C.1: A concave function.

If f is concave and is a differentiable function of many variables, then the Hessian matrix of second derivatives of: $f_{xy} \equiv \partial^2 f / (\partial x \partial y)$:

$$H = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix}, \quad (\text{C.4})$$

is negative semi-definite, i.e.: $x^T H x \leq 0^2$, for all x , which implies, among other things, that the diagonal elements are all negative and that the principal minor determinants of order k have the sign $:-, +, -, +, \dots$ for: $k = 1, 2, 3, 4, \dots$, i.e.:

$$f_{11} < 0, \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} > 0, \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix} < 0, \dots, (-1)^n |H| \geq 0, \quad (\text{C.5})$$

furthermore:

$$f(x) \leq f(y) + \frac{\partial f(y)}{\partial y} (x - y), \quad (\text{C.6})$$

where $\partial f(y) / \partial y$ is understood as a vector of first derivatives and the product is an inner product.

²where: $x_i = x_i^* - x_i^0$ the difference between any two points in the domain of f : $x = [x_1, \dots, x_n]$ and x^T is the transpose of x .

For a convex, differentiable function of several variables the Hessian matrix should be positive semi-definite, equally implying that: $x^T H x \geq 0$, for all x and:

$$f(x) \geq f(y) + \frac{\partial f(y)}{\partial y}(x - y). \quad (\text{C.7})$$

Most textbooks on dynamic optimisation and optimal control discuss convexity and concavity of functions in depth. The reader is, for example, referred to Kamien and Schwartz (1993, Appendix A, Section 3).

Appendix D

Partial derivatives

D.1 For optimal control problem A4/B3/C1

The partial derivative expressions in equations 5.121 to 5.129 are as follows, using the product rule:

$$\frac{d(f(x)g(x))}{dx} = \frac{df(x)}{dx}g(x) + f(x)\frac{dg(x)}{dx}, \quad (\text{D.1})$$

that is:

$$\begin{aligned} \frac{\partial \hat{V}_k(t)}{\partial V_l(t)} &= \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} (\check{Q}_i(t) \left(\frac{\partial D_{ij}(t)}{\partial V_l(t)} G_{ijmr}(t) + D_{ij}(t) \frac{\partial G_{ijmr}(t)}{\partial V_l(t)} \right)) \\ &= \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} \left(\frac{\check{Q}_i(t)}{V_l(t)} (\Omega_{2ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{3ijmr}(t)) \right), \end{aligned} \quad (\text{D.2})$$

$$\frac{\partial V_k(t)}{\partial V_l(t)} = \begin{cases} 1 & \text{if } k = l; \\ 0 & \text{if } k \neq l, \end{cases} \quad (\text{D.3})$$

$$\begin{aligned} \frac{\partial \hat{V}_k(t)}{\partial C_l^d(t)} &= - \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} \left(\frac{\check{Q}_i(t)}{C_l^d(t)} (\Omega_{2ij}(t) G_{ijmr}(t) + \right. \\ &\quad \left. D_{ij}(t) \Omega_{3ijmr}(t)) \right), \end{aligned} \quad (\text{D.4})$$

$$\begin{aligned} \frac{\partial \hat{V}_k(t)}{\partial I_l(t)} &= \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} \left(\frac{\check{Q}_i(t)}{I_l(t)} (\beta_5 \Omega_{2ij}(t) G_{ijmr}(t) + \right. \\ &\quad \left. \beta_5 D_{ij}(t) \Omega_{3ijmr}(t)) \right), \end{aligned} \quad (\text{D.5})$$

$$\frac{\partial \hat{V}_k(t)}{\partial U^v(t)} = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} \check{Q}_i(t) (\Omega_{4ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{5ijmr}(t)), \quad (\text{D.6})$$

$$\frac{\partial \hat{V}_k(t)}{\partial U_j^p(t)} = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} \check{Q}_i(t) (\Omega_{7ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{8ijmr}(t)), \quad (\text{D.7})$$

$$\frac{\partial R_k(t)}{\partial V_l(t)} = \Omega_9 \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} \Phi_{1m} \left(\frac{\check{Q}_i(t)}{V_l(t)} (\Omega_{2ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{3ijmr}(t)) \right), \quad (\text{D.8})$$

$$\frac{\partial R_k(t)}{\partial C_l^d(t)} = -\Omega_9 \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} \Phi_{1m} \left(\frac{\check{Q}_i(t)}{C_l^d(t)} (\Omega_{2ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{3ijmr}(t)) \right), \quad (\text{D.9})$$

$$\frac{\partial R_k(t)}{\partial I_l(t)} = \Omega_9 \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} \Phi_{1m} \left(\frac{\check{Q}_i(t)}{I_l(t)} (\beta_5 \Omega_{2ij}(t) G_{ijmr}(t) + \beta_5 D_{ij}(t) \Omega_{3ijmr}(t)) \right), \quad (\text{D.10})$$

$$\frac{\partial R_k(t)}{\partial U^v(t)} = \Omega_9 \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} \Phi_{1m} \check{Q}_i(t) (\Omega_{4ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{5ijmr}(t)), \quad (\text{D.11})$$

$$\frac{\partial R_k(t)}{\partial U_j^p(t)} = \Omega_9 \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} \Phi_{1m} \check{Q}_i(t) (\Omega_{7ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{8ijmr}(t)), \quad (\text{D.12})$$

$$\frac{\partial \hat{K}(t)}{\partial V_l(t)} = \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}^k} \theta_{1m} d_r \left(\frac{\check{Q}_i(t)}{V_l(t)} (\Omega_{2ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{3ijmr}(t)) \right), \quad (\text{D.13})$$

$$\frac{\partial \hat{K}(t)}{\partial C_l^d(t)} = -\alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \theta_{1m} d_r \left(\frac{\check{Q}_i(t)}{C_l^d(t)} (\Omega_{2ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{3ijmr}(t)) \right), \quad (\text{D.14})$$

$$\frac{\partial \hat{K}(t)}{\partial I_l(t)} = \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \theta_{1m} d_r \left(\frac{\check{Q}_i(t)}{I_l(t)} (\beta_5 \Omega_{2ij}(t) G_{ijmr}(t) + \beta_5 D_{ij}(t) \Omega_{3ijmr}(t)) \right), \quad (\text{D.15})$$

$$\frac{\partial \hat{K}(t)}{\partial U^v(t)} = \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \theta_{1m} d_r \check{Q}_i(t) (\Omega_{4ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{5ijmr}(t)), \quad (\text{D.16})$$

$$\frac{\partial \hat{K}(t)}{\partial U_j^p(t)} = \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \theta_{1m} d_r \check{Q}_i(t) (\Omega_{7ij}(t) G_{ijmr}(t) + D_{ij}(t) \Omega_{8ijmr}(t)), \quad (\text{D.17})$$

Using the triple product rule:

$$\frac{d(f(x)g(x)h(x))}{dx} = \frac{df(x)}{dx} g(x)h(x) + f(x) \frac{dg(x)}{dx} h(x) + f(x)g(x) \frac{dh(x)}{dx}, \quad (\text{D.18})$$

that is:

$$\frac{\partial \hat{P}(t)}{\partial V_l(t)} = \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \left(\frac{\check{Q}_i(t)}{V_l(t)} (\Omega_{2ij}(t) G_{ijmr}(t) \Omega_{10r}(t) + D_{ij}(t) \Omega_{3ijmr}(t) \Omega_{10r}(t) + D_{ij}(t) G_{ijmr}(t) \Omega_{11r}(t) \Omega_{12ijmr}(t)) \right), \quad (\text{D.19})$$

$$\frac{\partial \hat{P}(t)}{\partial C_l^d(t)} = -\alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \left(\frac{\check{Q}_i(t)}{C_l^d(t)} (\Omega_{2ij}(t) G_{ijmr}(t) \Omega_{10r}(t) + D_{ij}(t) \Omega_{3ijmr}(t) \Omega_{10r}(t) + D_{ij}(t) G_{ijmr}(t) \Omega_{11r}(t) \Omega_{12ijmr}(t)) \right), \quad (\text{D.20})$$

$$\begin{aligned} \frac{\partial \hat{P}(t)}{\partial I_l(t)} = & \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \left(\frac{\check{Q}_i(t)}{I_l(t)} (\beta_5 \Omega_{2ij}(t) G_{ijmr}(t) \Omega_{10r}(t) + \right. \\ & \left. \beta_5 D_{ij}(t) \Omega_{3ijmr}(t) \Omega_{10r}(t) + D_{ij}(t) G_{ijmr}(t) \Omega_{11r}(t) \Omega_{12ijmr}(t)) \right), \end{aligned} \quad (\text{D.21})$$

$$\begin{aligned} \frac{\partial \hat{P}(t)}{\partial U^v(t)} = & \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \check{Q}_i(t) (\Omega_{4ij}(t) G_{ijmr}(t) \Omega_{10r}(t) + \\ & D_{ij}(t) \Omega_{5ijmr}(t) \Omega_{10r}(t) + D_{ij}(t) G_{ijmr}(t) \Omega_{11r}(t) \Omega_{12ijmr}(t)), \end{aligned} \quad (\text{D.22})$$

$$\begin{aligned} \frac{\partial \hat{P}(t)}{\partial U_j^p(t)} = & \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \check{Q}_i(t) (\Omega_{7ij}(t) G_{ijmr}(t) \Omega_{10r}(t) + \\ & D_{ij}(t) \Omega_{8ijmr}(t) \Omega_{10r}(t) + D_{ij}(t) G_{ijmr}(t) \Omega_{11r}(t) \Omega_{12ijmr}(t)), \end{aligned} \quad (\text{D.23})$$

$$\begin{aligned} \frac{\partial \hat{E}(t)}{\partial V_l(t)} = & \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \sum_{l_r \in r} \theta_{1m} \left(\frac{\check{Q}_i(t)}{V_l(t)} (\Omega_{2ij}(t) G_{ijmr}(t) \Omega_{13mr}(t) + \right. \\ & \left. D_{ij}(t) \Omega_{3ijmr}(t) \Omega_{13mr}(t) + D_{ij}(t) G_{ijmr}(t) \Omega_{14mr}(t)) \right), \end{aligned} \quad (\text{D.24})$$

$$\begin{aligned} \frac{\partial \hat{E}(t)}{\partial C_l^d(t)} = & -\alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \sum_{l_r \in r} \theta_{1m} \left(\frac{\check{Q}_i(t)}{C_l^d(t)} (\Omega_{2ij}(t) G_{ijmr}(t) \Omega_{13mr}(t) + \right. \\ & \left. D_{ij}(t) \Omega_{3ijmr}(t) \Omega_{13mr}(t) + D_{ij}(t) G_{ijmr}(t) \Omega_{14mr}(t)) \right), \end{aligned} \quad (\text{D.25})$$

$$\begin{aligned} \frac{\partial \hat{E}(t)}{\partial I_l(t)} = & \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \sum_{l_r \in r} \theta_{1m} \left(\frac{\check{Q}_i(t)}{I_l(t)} (\beta_5 \Omega_{2ij}(t) G_{ijmr}(t) \Omega_{13mr}(t) + \right. \\ & \left. \beta_5 D_{ij}(t) \Omega_{3ijmr}(t) \Omega_{13mr}(t) + D_{ij}(t) G_{ijmr}(t) \Omega_{14mr}(t)) \right), \end{aligned} \quad (\text{D.26})$$

$$\begin{aligned} \frac{\partial \hat{E}(t)}{\partial U^v(t)} = & \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \sum_{l_r \in r} \theta_{1m} \check{Q}_i(t) (\Omega_{4ij}(t) G_{ijmr}(t) \Omega_{13mr}(t) + \\ & D_{ij}(t) \Omega_{5ijmr}(t) \Omega_{13mr}(t) + D_{ij}(t) G_{ijmr}(t) \Omega_{14mr}(t)), \end{aligned} \quad (\text{D.27})$$

$$\begin{aligned} \frac{\partial \hat{E}(t)}{\partial U_j^p(t)} = & \alpha_4 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{m \in \mathcal{M}} \sum_{r \in \mathcal{R}_{ij}} \sum_{l_r \in r} \theta_{1m} \check{Q}_i(t) (\Omega_{7ij}(t) G_{ijmr}(t) \Omega_{13mr}(t) + \\ & D_{ij}(t) \Omega_{8ijmr}(t) \Omega_{13mr}(t) + D_{ij}(t) G_{ijmr}(t) \Omega_{14mr}(t)), \end{aligned} \quad (\text{D.28})$$

$$\frac{\partial M_l(t)}{\partial U_l^m(t)} = \beta_7 \alpha_5 (I_l(t) - I_p) U_l^m(t)^{\beta_7 - 1}, \quad (\text{D.29})$$

$$\frac{\partial U_l^m(t)^{\max}}{\partial V_l(t)} = \frac{U_l^m(t)^{\max}}{\beta_7 \Omega_{11r}(t)} \gamma^2 \frac{\partial R_k(t)}{\partial V_l(t)}, \quad (\text{D.30})$$

$$\frac{\partial U_l^m(t)^{\max}}{\partial C_l^d(t)} = \frac{U_l^m(t)^{\max}}{\beta_7 \Omega_{11r}(t)} \gamma^2 \left(\frac{\partial R_k(t)}{\partial C_l^d(t)} + \frac{\alpha_6 (I_l(t) - I_p) U_l^c(t)}{C_l^{d^2}(t)} \right), \quad (\text{D.31})$$

$$\frac{\partial U_l^m(t)^{\max}}{\partial I_l(t)} = \frac{U_l^m(t)^{\max}}{\beta_7 \Omega_{11r}(t)} \left(1 + \gamma^2 \left(\frac{\partial R_k(t)}{\partial I_l(t)} + \xi_1 - \alpha_6 \frac{U_l^c(t)}{C_l^d(t)} \right) - \frac{\Omega_{11r}(t)}{\alpha_5 (I_l(t) - I_p)} \right), \quad (\text{D.32})$$

$$\frac{\partial U_l^m(t)^{\max}}{\partial U_l^c(t)} = - \frac{U_l^m(t)^{\max}}{\beta_7 \Omega_{11r}(t)} \frac{\gamma^2 \alpha_6 (I_l(t) - I_p)}{C_l^d(t)}, \quad (\text{D.33})$$

$$\Omega_{1ijmr}(t) = \begin{cases} \beta_{3m} \beta_1 \alpha_1 \tau_{lr}^0 \left(\frac{V_l(t)}{C_l^c(t)} \right)^{\beta_1} & \text{if } l \in r \{ \in \mathcal{R}_{ij} \}; \\ 0 & \text{if } l \notin r \{ \in \mathcal{R}_{ij} \}, \end{cases} \quad (\text{D.34})$$

$$\begin{aligned} \Omega_{2ij}(t) = & - \frac{b_1 \lambda_2 \exp(-\lambda_2 (c_{ij}^*(t) - c_{ij}^0))}{\sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} \exp(-\lambda_3 u_{ijm'r'}(t))} \\ & \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (\exp(-\lambda_3 u_{ijm'r'}(t)) \Omega_{1ijm'r'}(t)), \end{aligned} \quad (\text{D.35})$$

$$\Omega_{3ijmr}(t) = \lambda_1 G_{ijmr}(t) \begin{cases} \sum_{j' \in \mathcal{J}^i} \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (G_{ij'm'r'}(t) \Omega_{1ij'm'r'}(t)) & \text{if } l \notin r \in \mathcal{R}_{ij}; \\ \sum_{j' \in \mathcal{J}^i} \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (G_{ij'm'r'}(t) \Omega_{1ij'm'r'}(t)) - \Omega_{1ijmr}(t) & \text{if } l \in r \in \mathcal{R}_{ij}, \end{cases} \quad (\text{D.36})$$

$$\Omega_{4ij}(t) = - \frac{b_1 \lambda_2 \exp(-\lambda_2 (c_{ij}^*(t) - c_{ij}^0))}{\sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} \exp(-\lambda_3 u_{ijm'r'}(t))} \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (\exp(-\lambda_3 u_{ijm'r'}(t)) \beta_{4m'} \theta_{2m'} f_{r'}), \quad (\text{D.37})$$

$$\Omega_{5ijmr}(t) = \lambda_1 G_{ijmr}(t) \begin{cases} \sum_{j' \in \mathcal{J}^i} \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (G_{ij'm'r'}(t) \beta_{4m'} \theta_{2m'} f_{r'}) & \text{if } l \notin r \in \mathcal{R}_{ij}; \\ \sum_{j' \in \mathcal{J}^i} \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (G_{ij'm'r'}(t) \beta_{4m'} \theta_{2m'} f_{r'}) - \beta_{4m} \theta_{2m} d_r & \text{if } l \in r \in \mathcal{R}_{ij}, \end{cases} \quad (\text{D.38})$$

$$\Omega_{6ijmr}(t) = \begin{cases} \beta_{4m} \theta_{2m} & \text{if } j = j' \text{ in } r' \in \mathcal{R}_{ij}; \\ 0 & \text{if } j \neq j' \text{ in } r' \in \mathcal{R}_{ij}, \end{cases} \quad (\text{D.39})$$

$$\Omega_{7ij}(t) = - \frac{b_1 \lambda_2 \exp(-\lambda_2 c_{ij}^*(t) - c_{ij}^0)}{\sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} \exp(-\lambda_3 u_{ijm'r'}(t))} \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (\exp(-\lambda_3 u_{ijm'r'}(t)) \beta_{4m'} \theta_{2m'}), \quad (\text{D.40})$$

$$\Omega_{8ijmr}(t) = \lambda_1 G_{ijmr}(t) \begin{cases} \sum_{j' \in \mathcal{J}^i} \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (G_{ij'm'r'}(t) \Omega_{6ij'm'r'}(t)) & \text{if } l \notin r \in \mathcal{R}_{ij}; \\ \sum_{j' \in \mathcal{J}^i} \sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (G_{ij'm'r'}(t) \Omega_{6ij'm'r'}(t)) - \Omega_{6ijmr}(t) & \text{if } l \in r \in \mathcal{R}_{ij}, \end{cases} \quad (\text{D.41})$$

$$\Omega_9 = w \exp(vt) (1 + S_l)^{\beta_6}, \quad (\text{D.42})$$

$$\Omega_{10r}(t) = \frac{d_r}{\tau_r(t)}, \quad (\text{D.43})$$

$$\Omega_{11r}(t) = \left(\frac{-d_r}{\tau_r(t)^2} \right), \quad (\text{D.44})$$

$$\Omega_{12ijmr}(t) = \begin{cases} \beta_1 \alpha_1 \tau_{l_r}^0 \left(\frac{V_i(t)}{C_i^e(t)} \right)^{\beta_1} & \text{if } l \in r \{ \in \mathcal{R}_{ij} \}; \\ 0 & \text{if } l \notin r \{ \in \mathcal{R}_{ij} \}, \end{cases} \quad (\text{D.45})$$

$$\Omega_{13mr}(t) = \epsilon_{m|p} d_{l_r} \left(\frac{d_{l_r}}{\tau_l(t)} \right)^{-\beta_{8m|p}}, \quad (\text{D.46})$$

$$\Omega_{14mr}(t) = \frac{\Omega_{13mr}(t) \Omega_{12ijmr}(t) \beta_{8m|p}}{\tau_l(t)}, \quad (\text{D.47})$$

D.2 For optimal control problem A3/B1/C2

Here, only the partial derivatives for the functional J (including smoothing and penalty terms) of problem (A3/B1/C2) are given. The other derivatives can be directly taken, or be easily derived, from the results in previous paragraph.

$$\begin{aligned} \frac{\partial J(t)}{\partial V_i(t)} = & -\beta_9 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \left(\left(\tilde{Q}_i(t) \tilde{X}_j(t) \right)^{\frac{1}{2}} \exp(-\beta_9 c_{ij}^*(t)) \right. \\ & \left. \sum_{r \in \mathcal{R}_{ij}} \left(\frac{\exp(-\lambda_3 u_{ijmr})}{\sum_{r' \in \mathcal{R}_{ij}} \exp(-\lambda_3 u_{ijmr'})} \frac{\Omega_{1ijmr}(t)}{V_i(t)} \right) \right), \end{aligned} \quad (\text{D.48})$$

$$\begin{aligned} \frac{\partial J(t)}{\partial C_l^d(t)} = & \beta_9 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \left(\tilde{Q}_i(t) \tilde{X}_j(t) \right)^{\frac{1}{2}} \exp(-\beta_9 c_{ij}^*(t)) \\ & \sum_{r \in \mathcal{R}_{ij}} \left(\frac{\exp(-\lambda_3 u_{ijmr})}{\sum_{r' \in \mathcal{R}_{ij}} \exp(-\lambda_3 u_{ijmr'})} \frac{\Omega_{1ijmr}(t)}{C_l^d(t)} \right), \end{aligned} \quad (\text{D.49})$$

$$\frac{\partial J(t)}{\partial U_l^c(t)} = -2\alpha_{\tau_1} U_l^c(t), \quad (\text{D.50})$$

$$\frac{\partial J(t)}{\partial \tilde{E}(t)} = \begin{cases} -2\chi_1 (\tilde{E}(t) - \tilde{E}^*) & \text{if } (\tilde{E}(t) - \tilde{E}^*) > 0; \\ 0 & \text{if } (\tilde{E}(t) - \tilde{E}^*) \leq 0. \end{cases} \quad (\text{D.51})$$

D.3 For optimal control problem A4/B2/C3

Here, only the partial derivatives for the functional J (including smoothing and penalty terms) of problem (A4/B2/C3) are given. The other derivatives can be directly taken, or be easily derived, from the results in previous paragraph.

$$\frac{\partial J(t)}{\partial P(t)} = 1, \quad (\text{D.52})$$

$$\frac{\partial J(t)}{\partial U_l^t(t)} = -2\alpha_{71}U_l^t(t), \quad (\text{D.53})$$

$$\frac{\partial J(t)}{\partial E(t)} = \begin{cases} -2\chi_1(E(t) - E^*) & \text{if } (E(t) - E^*) > 0; \\ 0 & \text{if } (E(t) - E^*) \leq 0. \end{cases} \quad (\text{D.54})$$

D.4 For optimal control problem A4/B2/C4

Here, only the partial derivatives for the functional J (including smoothing and penalty terms) of problem (A4/B2/C4) are given. The other derivatives can be directly taken, or be easily derived, from the results in previous paragraph.

$$\begin{aligned} \frac{\partial J(t)}{\partial V_l(t)} = & -2 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{k \in \mathcal{K}} \left((\check{Q}_{i|k}(t) c_{ij|k}^*(t) - \bar{c}_{ij}^*(t_0)) \check{Q}_{i|k}(t) \right. \\ & \left. \frac{\sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (\exp(-\lambda_3 u_{ijm'r'|k}(t)) \Omega_{1ijm'r'|k}(t))}{\sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} \exp(-\lambda_3 u_{ijm'r'|k}(t))} \right), \end{aligned} \quad (\text{D.55})$$

$$\begin{aligned} \frac{\partial J(t)}{\partial C_l^d(t)} = & 2 \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}^i} \sum_{k \in \mathcal{K}} \left((\check{Q}_{i|k}(t) c_{ij|k}^*(t) - \bar{c}_{ij}^*(t_0)) \check{Q}_{i|k}(t) \right. \\ & \left. \frac{\sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} (\exp(-\lambda_3 u_{ijm'r'|k}(t)) \Omega_{1ijm'r'|k}(t))}{\sum_{m' \in \mathcal{M}} \sum_{r' \in \mathcal{R}_{ij}} \exp(-\lambda_3 u_{ijm'r'|k}(t))} \right), \end{aligned} \quad (\text{D.56})$$

$$\frac{\partial J(t)}{\partial U_l^t(t)} = -2\alpha_{71}U_l^t(t), \quad (\text{D.57})$$

$$\frac{\partial J(t)}{\partial E(t)} = \begin{cases} -2\chi_1(E(t) - E^*) & \text{if } (E(t) - E^*) > 0; \\ 0 & \text{if } (E(t) - E^*) \leq 0. \end{cases} \quad (\text{D.58})$$

Samenvatting

Dutch summary

DUURZAME ONTWIKKELING VAN STEDELIJK VERKEER EN VERVOER ALS DYNAMISCH OPTIMALISATIEPROBLEEM

Recente ontwikkelingen in stedelijk verkeer en vervoer dwingen beleidsmakers in toenemende mate bij de planning, het management en beheer en onderhoud van verkeersinfrastructuur rekening te houden met een groot aantal complexe en soms onderling tegenstrijdige belangen zoals filebestrijding, emissiereductie, gebruik van schaarse hulpbronnen, gelijkheid en bereikbaarheid.

De huidige oplossing voor de verkeersproblemen zelf en de veranderende en zwaarder wordende eisen gesteld aan verkeers- en vervoerbeleid, wordt vaak gezocht in uitbreiding van de capaciteit van aanwezige infrastructuur, het beter gebruiken van bestaande infrastructuur, het ontmoedigen danwel aanmoedigen van het gebruik van bepaalde vervoerswijzen of pogingen verplaatsingspatronen van mensen en goederen rechtstreeks te beïnvloeden, conform het principe van *voorspel-maak-beheers* (predict-provide-manage). De beslissingen daaromtrent zijn meestal gebaseerd op ervaring (zogenaamde best-practices), bestaande theorieën en methoden en technieken van vervoersplanning. Desondanks lijken het huidige verkeers- en vervoersysteem en de gebruikte methoden, technieken en modellen - zowel in ontwikkelde als in ontwikkelingslanden - niet noodzakelijkerwijs te voldoen aan de eisen van duurzame ontwikkeling. Een duurzaam verkeers- en vervoersysteem kan alleen worden bereikt wanneer gebruik wordt gemaakt van een nieuw paradigma voor duurzaam verkeer en vervoer in combinatie met een daarbij horend raamwerk voor analyse en ontwerp.

In dit proefschrift wordt een theoretisch kader in combinatie met een methodologie geschetst, waarin voorwaarden van duurzame ontwikkeling rechtstreeks zijn geïntegreerd in modellen voor verkeersbeleid en vervoersplanning, in bijzonder de vervoersvraagmodellen. Met de opgedane kennis en de uitkomsten van dit onderzoek zal het mogelijk zijn efficiënter en effectiever gebruik te maken van beschikbare en betaalbare schaarse hulpbronnen in het belang van de ontwikkeling van het verkeers- en vervoersysteem zelf.

Met het paradigma van duurzaam verkeer en vervoer wordt een verkeers- en

vervoersysteem aangeduid binnen randvoorwaarden van milieu, energie, verkeersveiligheid en beschikbare budgetten. Dit paradigma is gebaseerd op de mondiale boodschap van wetenschappers en politici dat ongebreidelde groei op onze planeet zal leiden tot overmatig en onomkeerbaar gebruik van milieukundige en ecologische bronnen en uiteindelijk zelfs tot onze mogelijke ondergang. Het paradigma van een duurzame transportontwikkeling sluit hierbij aan en probeert een veelomvattend beslissingsondersteunend systeem te promoten, waarin ontwikkelingen in verkeer en vervoer in termen van mobiliteit, bereikbaarheid en verplaatsingen zich afspelen binnen de eerder genoemde randvoorwaarden van milieu, energie en beschikbare budgetten. Het systeem anticipeert derhalve op beheerst gebruik van schaarse bronnen, terwijl het zich (gestuurd) ontwikkelt in termen van kwaliteit en kwantiteit van doorstroming.

De afgelopen decennia zijn vervoersvraagmodellen ter ondersteuning van het complexe proces van vervoersplanning geëvolueerd van een simpele heuristiek tot een geavanceerd sequentieel vierstaps-verkeersmodel met deelmodellen voor ritgeneratie, ritdistributie, vervoerswijzekeuze en toedeling. Dit geavanceerde model is desondanks nog regelmatig bron van kritiek, met name in relatie tot het thema duurzame ontwikkeling en verkeer en vervoer. Zo is er kritiek op het sequentiële en statische karakter van het vierstaps-verkeersmodel, waardoor nauwelijks sprake kan zijn van een (simultane) terugkoppeling naar beslissingen in voorgaande stappen van dit model. Daarnaast bestaat er kritiek ten aanzien van de inelasticiteit van het ritgeneratie deelmodel voor veranderingen in bereikbaarheid, waardoor het verkeersmodel ongevoelig is voor - de algemeen aangenomen - aanzuigende werking van nieuwe infrastructuur. Ook het eerder genoemde typische gebruik van het verkeersmodel conform het principe van voorspel-maak-beheers, een vraagvolgende benadering, stuit vaak op kritiek. Limieten aan schaarse bronnen vereisen immers een aanbodgestuurde benadering conform het principe van maak-beheers-voorspel. Een verkeersmodel moet derhalve al redenerende vanuit een expliciet verkeersbeleidsdoel en kennis over beschikbare schaarse bronnen, komen tot een set aan verkeer- en vervoer gerelateerde maatregelen.

Naast deze punten van kritiek, kan ook worden gesteld dat het traditionele verkeersmodel minder bruikbaar is voor beleidsmakers als dit model geen rekening houdt met de continue verandering in het verplaatsingsgedrag van de gebruikers zelf (de reizigers), de prestatie van het verkeers- en vervoerssysteem (in termen van kwaliteit en kwantiteit van doorstroming) en de complexe en onderling afhankelijke doelen van betrokken beleidsmakers. Het zijn namelijk niet de individuele wensen en beslissingen van de betrokken gebruikers en beleidsmakers die bepalend moeten zijn voor de te nemen maatregelen, maar de onderling samenhangende toestanden van hun individuele systemen. Dynamische optimalisatie is bij uitstek geschikt om tegemoet te komen aan bovengenoemde kritiek. Hierbij worden tegelijkertijd de volgorde en het tijdstip van te nemen verkeersgerelateerde maatregelen bepaald, om zo een vooraf geëxpliciteerd verkeersbeleidsdoel te kunnen verwezenlijken op een vooraf vastgestelde termijn.

Dit lijkt dan ook een logische vervolgstap in de ontwikkeling van verkeersmodellen.

In dit onderzoek is een dergelijk verkeersmodel ontwikkeld, besproken en geïmplementeerd, gebaseerd op een conceptualisatie en karakterisering van duurzame transportontwikkeling als optimalisatieprobleem. De bouwstenen van het traditionele vierstaps-verkeersmodel worden zeer bruikbaar verondersteld voor de vraagstelling in dit onderzoek en worden dus ook als zodanig gebruikt bij het opstellen van het dynamische model. Een proactieve benadering van verkeersproblemen conform het principe van maak-beheers-voorspel ('voorkom'), in tegenstelling tot dat van voorspel-maak-beheers, is het resultaat, waardoor een wezenlijke bijdrage geleverd kan worden in het complexe (politieke) besluitvormingsproces rondom duurzaam verkeer en vervoer.

Het dynamische model maakt gebruik van een relaxatieformulering, waarin de toestandsvariabelen zich geleidelijk aanpassen aan het gewenste evenwicht, afhankelijk van de snelheid van aanpassing, die kan verschillen per type toestandsvariabele. In het transportmodel wordt een aantal standaard toestandsvariabelen, - zoals verkeersvraag (per link in het netwerk), infrastructuraanbod (en daarmee de effectieve capaciteit), het aantal gereden kilometers in het netwerk, de reizigersdoorstroming door het netwerk en de totale emissie van schadelijke stoffen in het netwerk - gemodelleerd middels gewone differentiaalvergelijkingen. De verkeersvraag wordt bepaald vanuit sociaaleconomische variabelen in de zones van het studiegebied, gebruikmakend van een discrete-keuze model, analoog aan het zwaartekrachtmodel, en is in omvang mede afhankelijk van de verkeersprestatie in het netwerk. De kwaliteit van het wegdek van een link, en daarmee de effectieve capaciteit, neemt autonoom af, danwel door slijtage ten gevolge van aslasten van het verkeer. De evenwichtswerking tussen verkeersvraag en infrastructuraanbod wordt continu verstoord door bestemmings-, vervoerswijze- en routekeuzegedrag van reizigers, maar ook door de aanzuigende werking van verkeersinfrastructuur. De vervoersplanner staan verschillende beslissingsvariabelen ter beschikking, die dit evenwicht proberen te beïnvloeden of te sturen. Hij kan zo aanbodzijdegerelateerde beslissingsvariabelen zoals capaciteitsuitbreiding of onderhoud, maar ook vraagzijdegerelateerde beslissingsvariabelen als beprijzing inzetten. Gestuurd door een gekwantificeerd verkeersbeleidsdoel - zoals congestieminimalisatie, bereikbaarheidsmaximalisatie, reizigersdoorstromingsmaximalisatie of gelijkheidsmaximalisatie - maar ook door beperkingen opgelegd aan de waarde van de beslissingsvariabele of die van een toestandsvariabele - zoals totale emissies - wordt nu de regelstrategie (in termen van volgorde en het moment van inzetten van de beslissingsvariabelen) bepaald.

Het nieuwe verkeersmodel kan in zijn huidige vorm direct worden gebruikt voor de modellering van strategische netwerken met een beperkt aantal (geaggregeerde) zones en (geaggregeerde) links. Uitgewerkt als randwaardeprobleem in continue-tijd met het Pontryagin Maximum Principe, produceert het model regelstrategieën om een duurzaam en ontwikkeld verkeers- en vervoersysteem te

bereiken, waarbij tegelijkertijd allerlei informatie over de toestandsvariabelen en de bijbehorende schaduwrijzen (de zogenaamde costates) wordt gegenereerd. Het dynamische model is desondanks nog niet volledig geïmplementeerd met een 'complete' dataset, waardoor thans geen algemeen geldende beleidsconclusies kunnen worden getrokken. De belangrijkste reden hiervoor is de gebleken tekortkoming in robuustheid van de gekozen optimalisatiemethode, met name in relatie tot de interactie van beslissingsvariabelen en bij niet-lineaire beslissingsvariabelen. Desondanks kan zowel uit deze eerste ontwikkelingsstap van het complexe model, als op grond van de diverse cases opgemaakt worden dat het haalbaar zal zijn een volledig bruikbaar model te implementeren, parameters te schatten en het model toe te passen in een strategische vervoersplanningsstudie.

Bibliography

- Akinyemi, E. O. (1997). A total performance approach to traffic management for sustainable development of communities in third-world cities, *Resource Papers for the 1997 ITE International Conference*, Tampa, USA, pp. 191–209.
- Akinyemi, E. O. (1998). Engineering and urban mobility enhancement in developing countries, *CODATU VIII Conference*, Cape Town, South Africa, pp. 715–719.
- Akinyemi, E. O. and Zuidgeest, M. H. P. (2000). Sustainable development & transportation: past experiences and future challenges, *World Transport Policy and Practice* **6**(1): 31–39.
- Akinyemi, E. O. and Zuidgeest, M. H. P. (2002). Managing transportation infrastructure for sustainable development, *Computer-Aided Civil and Infrastructure Engineering* **17**(3): 148–161.
- Al-Azzawi, M. (2001). Developing a ‘rapid’ strategic land-use and transport policy model using EMME/2, *10th European EMME/2 Users Group Conference*, Thessaloniki, Greece.
- Al-Deek, H., Wayson, R. and Radwan, A. (1995). Methodology for evaluating ATIS impacts on air quality, *Journal of Transportation Engineering* **July/August 1995**: 376–384.
- Anderson, J. E. (2003). *Public Policymaking*, fifth edn, Houghton Mifflin Company, Boston, USA.
- Andersson, J. (2001). *Multiobjective Optimization in Engineering Design: Applications to Fluid Power Systems*, PhD thesis, Linköping Studies in Science and Technology No. 675, University of Linköping, Linköping, Sweden.
- Asian Development Bank (2003). Appendix on adverse health and environmental effects from vehicle emissions, *Policy Guidelines for Reducing Vehicle Emissions in Asia*, Asian Development Bank, Manila, The Philippines.
- Banister, D. (2002). *Transport Planning*, Transport, Development and Sustainability, second edn, Sponn Press, London, England.

- Banister, D. and Berechman, J. (2000). *Transport Investment and Economic Development*, UCL Press, London, England.
- Barkenbus, J. (1998). Expertise and the policy cycle, *Report NCEDR/99-04*, University of Tennessee: Energy, Environment, and Resources Center (NCEDR), Knoxville, USA.
- Barten, H. (2004). *Optimalisatie in de Vervoersplanning*, Master's thesis, University of Twente, Enschede, The Netherlands. [In Dutch].
- Bates, J. (2001). History of demand modelling, in D. A. Hensher and K. J. Button (eds), *Handbook of Transport Modelling*, Vol. 1 of *Handbooks in Transport*, Pergamon, Amsterdam, The Netherlands, chapter 2, pp. 11–33.
- Becker, E., Jahn, T., Stiess, I. and Wehling, P. (1997). Sustainability: A cross-disciplinary concept for social transformations, *MOST Policy Papers 6*, United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris, France.
- Bell, M. G. H. and Allsop, R. E. (1998). Introduction, in M. G. H. Bell (ed.), *Transportation Networks: Recent Methodological Advances*, Selected proceedings of the 4th Euro transportation meeting, Elsevier Science & Technology, Oxford, England.
- Bell, M. G. H. and Iida, Y. (1999). *Transportation Network Analysis*, John Wiley & Sons, Chichester, England.
- Ben-Akiva, M. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Applications to Travel Demand*, The MIT Press, Cambridge, Massachusetts, USA.
- Bertrand, R. and Epenoy, R. (2002). New smoothing techniques for solving bang-bang optimal control problems - numerical results and statistical interpretation, *Optimal Control Applications and Methods* **23**: 171–197.
- Bierman, M. O. (2004). *Induced Travel: Relating Accessibility and Trip Frequency*, Master's thesis, University of Twente, Enschede, The Netherlands.
- Bluemel, H. (1999). The berlin urban traffic master plan: new approaches to the integration of transport, land use and environmental planning, *Clean Air Initiative in Latin American Cities. Clean Air Workshop for the Mexico City Metropolitan Area*, Mexico City, Mexico.
- Bovy, P. H. L. and Van der Zijpp, N. J. (1999). Transportation modelling, *Lecture Note CTvk4800*, Delft University of Technology, Delft, The Netherlands.
- Bryson, A. E. and Ho, Y. (1975). *Applied Optimal Control: Optimization, Estimation and Control*, Taylor & Francis, USA.

- BTE (1996). Roads, vehicle performance and greenhouse: costs and emission benefits of smoother highways, *Working Paper 32*, Bureau of Transport and Communications Economics, Australia.
- BTE (1998). Urban transport models: A review, *Working Paper 39*, Bureau of Transport and Communications Economics, Australia.
- Bureau of Public Roads (1964). *Traffic Assignment Manual*, Department of Commerce, Urban Planning Division, Washington D.C., USA.
- Büttler, H. and Shortreed, J. H. (1978). Investment planning of a road link, *Transportation Research* **12**: 357–366.
- Camagni, R. (1998). Sustainable urban development: definition and reasons for a research programme, *International Journal on Environment and Pollution* **10**(1): 6–27.
- Camagni, R., Capello, R. and Nijkamp, P. (1998). Towards sustainable city policy: an economy-environment technology nexus, *Ecological Economics* **24**: 103–118.
- Cervero, R. (2002). Induced travel demand: research design, empirical evidence, and normative policies, *Journal of Planning Literature* **17**(1): 3–20.
- Chiou, S. W. (1999). Optimization of area traffic control for equilibrium network flows, *Transportation Science* **33**(3): 279–289.
- Colucci-Rios, B. and Sinha, K. (1985). Optimal pavement management approach using roughness measurements, *Transportation Research Record* **1048**: 14–22.
- Constantin, I. and Florian, M. (1995). Optimizing frequencies in a transit network: a nonlinear bi-level programming approach, *Transactions in Operational Research* **2**(2): 149–164.
- Crozet, Y. and Marlot, G. (2001). Congestion and road pricing: where is the “bug”?, *Proceedings of the 9th World Conference on Transport Research*, Seoul, South Korea.
- CST (1997). Definition and vision of sustainable transportation, *Technical report*, The Centre for Sustainable Transportation (CST), Toronto, Canada.
- Dadebo, S. A., McAuley, K. B. and McLellan, P. J. (1998). On the computation of optimal singular and bang-bang controls, *Optimal Control Applications and Methods* **19**: 287–297.
- Dahlgren, J. (2003). The benefits of accommodating latent demand, *Transportation Research Board (TRB) 2003 Annual Meeting CD-ROM*, Washington D.C., USA.
- Daly, H. E. (1991). *Steady-State Economics*, second edn, Island Press, Washington D.C., USA.

- De Borger, B. and Wouters, S. (1998). Transport externalities and optimal pricing and supply decisions in urban transportation, *Regional Science and Urban Economics* **28**(2): 163–197.
- De Cea, J. and Fernandez, E. (2001). Transit-assignment models, in D. A. Hensher and K. J. Button (eds), *Handbook of Transport Modelling*, Vol. 1 of *Handbooks in Transport*, Pergamon, Amsterdam, The Netherlands, chapter 30.
- De la Barra, T. (1989). *Integrated Land Use and Transport: Decision Chains and Hierarchies*, Cambridge University Press, Cambridge, New York, USA.
- De la Barra, T. (n.d.). Improved logit formulations for integrated land use, transport and environmental models, *Modelistica Systems and Planning*.
*<http://www.modelistica.com/papers/suecia/suecia.htm>
- DETR (2002). Social exclusion and the provision and availability of public transport, *Technical report*, Department of the Environment, Transport and the Regions (DETR), London, England.
- Donaghy, K. P. and Schintler, L. A. (1994). Optimal policies for reducing congestion of major commuting arteries in a growing urbanized area, *Journal of Planning and Research* **13**: 189–198.
- Donaghy, K. P. and Schintler, L. A. (1998). Managing congestion, pollution, and pavement conditions in a dynamic transportation network model, *Transportation Research* **3D**(2): 59–80.
- ECMT (1995). *Urban Travel and Sustainable Development*, European Conference of Ministers of Transport (ECMT), Paris, France.
- English dictionary (2000). *The American Heritage Dictionary of the English Language*, Houghton Mifflin Publishers, Massachusetts, USA.
- ESCAP-AITD (2001). Sustainable transport pricing and charges: Principles and issues, *Technical report*, United Nations - Economic and Social Commission for Asia and the Pacific (UN-ESCAP), Asian Institute of Transport Development (AITD), Bangkok, Thailand.
- European Commission (1998). *OPTIMA: Optimisation of Policies for Transport Integration in Metropolitan Areas*, Office for Official Publications of the European Communities, Luxembourg.
- European Commission (2000). EU transport in figures, *Technical report*, European Commission Directorate-General for Energy and Transport, Brussels.
- Eurostat (2001). Panorama of transport: Statistical overview of road, rail, inland waterways and air transport in the European Union (data 1970 - 1999), *Technical report*, Eurostat, European Union (EU), Luxembourg.

- Ferrari, P. (1995). Road pricing and network equilibrium, *Transportation Research* **29B**(5): 357–372.
- Ferrari, P. (1999). A model of urban transport management, *Transportation Research* **33B**(1): 43–61.
- Friesz, T. L. (1985). Transportation network equilibrium, design and aggregation: key developments and research opportunities, *Transportation Research* **19A**(5/6): 413–427.
- Friesz, T. L. and Bernstein, D. (2001). Analytical dynamic traffic assignment models, in D. A. Hensher and K. J. Button (eds), *Handbook of Transport Modelling*, Vol. 1 of *Handbooks in Transport*, Pergamon, Amsterdam, The Netherlands, chapter 11, pp. 181–195.
- Friesz, T. L. and Fernandez, J. E. (1979). A model of optimal transport maintenance with demand responsiveness, *Transportation Research* **13B**(4): 317–339.
- Friesz, T. L. and Shah, S. (1998). Foundations of a theory of disequilibrium network design, in M. G. H. Bell (ed.), *Transportation Networks: Recent Methodological Advances*, Selected Proceedings of the 4th Euro Transportation Meeting, Elsevier Science & Technology, Oxford, England.
- Gandolfo, G. (1993). Continuous-time econometrics has come of age, in G. Gandolfo (ed.), *Continuous Time Econometrics: Theory and Applications*, Chapman & Hall, London, England, chapter 1, pp. 1–11.
- Gandolfo, G. (1997). *Economic Dynamics*, Springer, Berlin, Germany.
- Gauthier, H. (1970). Geography, transportation and regional development, *Economic Geography* **46**: 612–619.
- Geurs, K. T. and Ritsema van Eck, J. R. (2001). Accessibility measures: Review and applications, *Report RIVM 408505 006*, RIVM National Institute of Public Health and the Environment, Bilthoven, The Netherlands.
- Goodwin, P. B. (1992). A review of new demand elasticities with special reference to short and long run effects of price changes, *Journal of Transport Economics and Policy* **26**(2): 155–169.
- Goodwin, P. B. (1996). Empirical evidence on induced traffic, *Transportation* **23**: 35–54.
- Goodwin, P. B. (1999). Transformation of transport policy in Great Britain, *Transportation Research* **33A**(7-8): 655–669.
- Goodwin, P., Kitamura, R. and Meurs, H. (1990). Some principles of dynamic analysis of travel behaviour, in P. Jones (ed.), *Developments in Dynamic and Activity-based Approaches to Travel Analysis*, Aldershot, Avebury, England, pp. 56–73.

- Goudappel Coffeng (2003). PROMILspatial, *Computer program*, Goudappel Coffeng consultants, Deventer, The Netherlands.
- Gudmundsson, H. (2001). Indicators for performance measures for transportation, environment and sustainability in North America, *German Marshall Fund Fellowship report 148*, National Environmental Research Institute, Roskilde, Denmark.
- Gudmundsson, H. (2003). Concepts of sustainable transport: making sense or not?, *PhD course: Sustainable Transport Futures*, Roskilde, Denmark.
- Gudmundsson, H. and Höjer, M. (1996). Sustainable development principles and their implications for transport, *Ecological Economics* **19**: 269–282.
- Haas, R. C. G. and Hudson, W. R. (1978). *Pavement Management Systems*, Krieger Publishing Company, Florida, USA.
- Hansen, W. G. (1959). How accessibility shapes land use, *Journal of American Institute of Planners* **25**(2): 72–76.
- Hau, T. D. (1992). Economic fundamentals of road pricing: A diagrammatic analysis, *World Bank Policy Research Working Paper Series WPS 1070*, The World Bank, Washington D.C., USA.
- Hilling, D. (1996). *Transport and Developing Countries*, Routledge, London, England.
- Hillman, M. and Adams, J. (1992). Safer driving - safer for whom?, *Conference on Eurosafe: Safer Driving in Europe*, The association of London Borough road safety officers, London, England.
- Hinloopen, E., Nijkamp, P. and Rietveld, P. (1983). The regime method: a new multi-criteria technique, in P. Hansen (ed.), *Essays and Surveys on Multiple Criteria Decision Making*, Springer Verlag, Berlin, Germany, pp. 146–155.
- Höjer, M. (1998). Transport telematics in urban systems: a backcasting Delphi study, *Transportation Research* **3D**(6): 445–463.
- Höjer, M. and Mattsson, L. (2000). Determinism and backcasting in future studies, *Futures* **32**: 613–634.
- Huapu, L. and Peng, Z. (2001). A study on urban transport planning model integrating land-use mode and environment capacity based on sustainable development, *Proceedings of the 9th World Conference on Transport Research*, Seoul, South Korea.
- ICLEI (1996). *The Local Agenda 21 Planning Guide: An Introduction to Sustainable Development Planning*, International Council on Local Environment Initiatives (ICLEI), Toronto, Canada.

- Immers, L. H. and Oosterbaan, C. G. J. (1991). A model for the calculation of environment-friendly traffic flows in urban networks, *Proceedings of the 70th Annual TRB Meeting*, Washington D.C., USA.
- Intriligator, M. D. and Sheshinski, E. (1986). Toward a theory of planning, in W. P. Heller, R. M. Starr and D. A. Starett (eds), *Social Choice and Public Decision Making*, Vol. 1 of *Essays in Honour of Kenneth J. Arrow*, Cambridge University Press, England, chapter II-7.
- Jansen, G. R. M. and Bovy, P. H. L. (1982). The effects of zone size and network detail on all-or-nothing and equilibrium assignment outcomes, *Traffic Engineering and Control* **23**(6): 311–317.
- Jayakrishnan, R., Tsai, W. K. and Chen, A. (1995). A dynamic traffic assignment model with traffic flow relationships, *Transportation Research* **3C**(1): 51–72.
- Kamien, M. I. and Schwartz, N. L. (1993). *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, North-Holland, New York, USA.
- Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing: Estimation Theory*, Vol. 1, Prentice Hall PTR, New Jersey, USA.
- Kierzenka, J. and Shampine, L. F. (2001). A BVP solver based on residual control and the Matlab PSE, *ACM Transactions on Mathematical Software* **27**(3): 299–316.
- Kim, K. (1998). *A Transportation Planning Model for State Highway Management: A Decision Support System Methodology to Achieve Sustainable Development*, PhD thesis, Virginia Polytechnic Institute and State University, Virginia, USA.
- Kim, T. J. and Hoskote, N. G. (1983). Estimating mobile source pollutant emission: methodological comparison and planning implications, *Environmental Monitoring and Assessment* **3**: 1 – 12.
- Knoflacher, H., Pfaffenbichler, P. C. and Emberger, G. (2000). A strategic transport model-based tool to support urban decision making processes, in J. C. Mangin and M. Miramond (eds), *2nd International Conference on Decision Making in Urban and Civil Engineering*, Lyon, France.
- Lasswell, H. D. (1951). The policy orientation, in D. Lerner and H. D. Lasswell (eds), *The Policy Sciences*, Stanford University Press, Palo Alto, USA.
- Lee, D. B. (1994). Retrospective on large-scale urban models, *Journal of the American Planning Association* **Winter 1994**(1): 35–40.
- Lensink, S. M. (2002). Applying optimal control to minimize energy use due to road infrastructure expansion, *Interim Report IR-02-071*, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

- Léonard, D. and Van Long, N. (1992). *Optimal Control Theory and Static Optimization in Economics*, Cambridge University Press, Cambridge, England.
- Lindquist, E. (1998). Moving towards sustainability: transforming a comprehensive land use and transportation plan, *Transportation Research Record* **1617**: 1–9.
- Litman, T. A. (n.d.a). Reinventing transportation: Exploring the paradigm shift needed to reconcile transportation and sustainability objectives, *Victoria Transport Policy Institute* . *<http://www.vtpi.org/>
- Litman, T. A. (n.d.b). Sustainable transportation indicators, *Victoria Transport Policy Institute* . *<http://www.vtpi.org/>
- Litman, T. A. and Burwell, D. (n.d.). Issues in sustainable transportation, *Victoria Transport Policy Institute* . *<http://www.vtpi.org/>
- Malone, K. M. (2001). The Scenario Explorer for passenger transport: a strategic model for long term travel demand forecasting, *Technical report*, TNO-INRO, Delft, The Netherlands.
- Manheim, M. L. (1979). *Fundamentals of Transportation System Analysis*, MIT Press, Cambridge, Massachusetts.
- Martinez-Alier, J. (n.d.). Ecological - economics, *International encyclopedia of the social and behavioral sciences*, Elsevier Science & Technology. article 4.9, number 9. *<http://www.iesbs.com/>
- May, A. D., Matthews, B. and Jarvi-Nykanen, T. (2001). Decision making requirements for the formulation of sustainable urban land use - transport strategies, *Proceedings of the 9th World Conference on Transport Research*, Seoul, South Korea.
- Meadows, D. H., Meadows, D. L. and Randers, J. (1992). *Beyond the Limits: Confronting Global Collapse, Envisioning a Sustainable Future*, Earthscan publications, London, England.
- Meadows, D. H., Meadows, D. L., Randers, J. and Behrens, W. W. (1972). *The Limits to Growth*, Universe books, New York, USA.
- Meng, Q. and Yang, M. (2002). Benefit distribution and equity in road network design, *Transportation Research* **36B**: 19–35.
- Meyer, M. D. and Miller, E. J. (2001). *Urban Transportation Planning*, McGraw Hill, New York, USA.
- Miller, J. S. and Demetsky, M. J. (1999). Reversing direction of transportation planning process, *Journal of Transportation Engineering* **May/June 1999**: 231–237.

- Ministerie van de Vlaamse Gemeenschap (2001). Naar een duurzame mobiliteit in Vlaanderen, *Mobiliteitsplan Vlaanderen*, Departement Leefmilieu en Infrastructuur, Brussels, Flanders, Belgium. [In Dutch].
- Morgan, M. G. and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, Cambridge, England.
- Morris, J. M., Dumble, P. L. and Wigan, M. R. (1979). Accessibility indicators for transport planning, *Transportation Research* **13A**: 91–109.
- Nagurney, A. (2000). *Sustainable Transportation Networks*, Edward Elgar Publishing, Inc., USA.
- Naidu, D. S. (2003). *Optimal Control Systems*, CRC Press, Florida, USA.
- NCHRP (1997). Quantifying congestion, *NCHRP report 398*, National Cooperative Highway Research Program (NCHRP), NAP, Washington D.C., USA.
- Newman, P. and Kenworthy, J. (1999). *Sustainability and Cities: Overcoming Automobile Dependence*, Island Press, Washington D.C., USA.
- Nijkamp, P. and Opschoor, H. (1995). Urban environmental sustainability: critical issues and policy measures in a third world context, in Chatterjee (ed.), *Urban Policies in Third World Countries*, MacMillan, New York, USA.
- Nijkamp, P. and Reggiani, A. (1988). Analysis of dynamic spatial interaction models by means of optimal control, *Geographical Analysis* **20**(1): 18–30.
- NIST-SEMATECH (n.d.). *e-Handbook of Statistical Methods*, National Institute of Standards and Technology - SEMiconductor MANufacturing TECHNOLOGY, USA. *<http://www.itl.nist.gov/div898/handbook/>
- Noland, R. B. (2001). Relationships between highway capacity and induced vehicle travel, *Transportation Research* **35A**: 47–72.
- Noland, R. B. and Lem, L. L. (2002). A review of the evidence for induced travel and changes in transportation and environmental policy in the US and the UK, *Transportation Research* **7D**: 1–26.
- Odoki, J. B., Kerali, H. R. and Santorini, F. (2001). An integrated model for quantifying accessibility-benefits in developing countries, *Transportation Research* **35A**: 601–623.
- Oppenheim, N. (1995). *Urban Travel Demand Modeling: From Individual Choices to General Equilibrium*, John Wiley & Sons, New-York, USA.
- Ortúzar, J., Martínez, F. J. and Varela, F. J. (2000). Stated preference modeling accessibility, *International Planning Studies* **5**: 65–85.

- Ortúzar, J. and Willumsen, L. G. (2001). *Modelling Transport*, third edn, John Wiley & Sons, Chichester, England.
- Oude Moleman, T. (2001). *Sustainable Performance - Measuring Sustainability of an Urban Transport System*, Master's thesis, University of Twente, Enschede, The Netherlands.
- Owen, W. (1987). *Transportation and World Development*, Hutchinson, London, England.
- Paterson, W. D. O. (1987). *Road Deterioration and Maintenance Effects: Models for Planning and Management*, The Johns Hopkins University Press, Baltimore, USA.
- Pearce, D. W. and Turner, R. K. (1990). *Economics of Natural Resources and the Environment*, Harvester Wheatsheaf, London, England.
- Peters, D. (2002). Life in times of the traffic jam: why smart cities are demotorising, *Sustainable Development International* **2002**(3): 157–161.
- Pfaffenbichler, P. C. (2003). *The Strategic, Dynamic and Integrated Urban Land Use and Transport Model MARS (Metropolitan Activity Relocation Simulator)*, PhD thesis, University of Vienna, Austria.
- Pfaffenbichler, P. C. and Shepherd, S. P. (2002). A dynamic model to appraise strategic land-use and transport policies, *European Journal of Transport and Infrastructure Research* **2**(3/4): 255–283.
- Pigou, A. C. (1932). *The Economics of Welfare*, 4 edn, Macmillan and Co., London, England.
- Plaut, P. O. and Shmueli, D. F. (2000). Sustainable transport: a comparative analysis of Israel, the Netherlands and the United Kingdom, *World Transport Policy and Practice* **6**(1): 40–58.
- Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V. and Mischenko, E. F. (1962). *The Mathematical Theory of Optimal Processes*, John Wiley & Sons, New York, USA.
- Prakash, A. B., D'A. Oliver, E. H. and Balcombe, K. (2001). Does building new roads really create extra traffic? Some new evidence, *Applied Economics* **33**: 1579–1585.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (eds) (1989). *Numerical Recipes in Pascal: The Art of Scientific Computing*, Cambridge University Press, Cambridge, England.
- Replogle, M. (1991). Sustainability: a vital concept for transportation planning and development, *Journal of Advanced Transportation* **25**(1): 3–18.
- Roozenburg, N. F. M. and Eekels, J. (1998). *Produktontwerpen, Structuur en Methoden*, Malberg, Utrecht, The Netherlands. [In Dutch].

- SACTRA (1994). Trunk roads and the generation of traffic, *Technical report*, Department of transport, Standing Advisory Committee on Trunk Road Assessment (SACTRA), London, England.
- Sales Filho, L. H. (1998). The accessibility matrix: a new approach for evaluating urban transportation networks, *Proceedings of the 8th World Conference on Transport Research (WCTR)*, Antwerp, Belgium.
- Schwartz, A. L. (1996). *Theory and Implementation of Numerical Methods Based on Runge-Kutta Integration for Solving Optimal Control Problems*, PhD thesis, University of California at Berkeley, USA.
- Schwela, D. (2002). Air quality management, in GTZ (ed.), *Sustainable transport: A sourcebook for policy-makers in developing cities*, Vol. 5a, Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ), Eschborn, Germany.
- Schwela, D. and Zali, O. (eds) (1999). *Urban Traffic Pollution*, E&FN Spon, London, England.
- Seierstad, A. and Sydsæter, K. (1987). *Optimal Control Theory with Economic Applications*, Elsevier Science & Technology, Amsterdam, The Netherlands.
- Shampine, L. F., Kierzenka, J. and Reichelt, M. W. (2000). Solving boundary value problems for ordinary differential equations in Matlab with `bvp4c`, *Matlab documentation*. *ftp://ftp.mathworks.com/pub/doc/papers/bvp/
- Sheffi, Y. (1985). *Urban Transportation Networks*, Prentice-Hall, Englewood Cliffs, USA.
- Shrestha, R. M., Anandarajah, G., Adhikari, S., Jiang, K. and Songli, Z. (2005). Energy and environmental implications of NO_x emission reduction from the transport sector of Beijing: a least-cost planning analysis, *Transportation Research* **10D**: 1–11.
- Stackelberg, H. V. (1934). *Marktform und Gleichgewicht*, Julius Springer, Vienna, Austria. [In German].
- Steenbrink, P. A. (1974). *Optimization of Transport Networks*, John Wiley & Sons, London, England.
- Tagore, M. R. and Sikdar, P. K. (1995). A new accessibility measure accounting mobility parameters, *Selected Proceedings of the 7th World Conference on Transport Research (WCTR)*, Vol. 1, Sydney, Australia, pp. 305–315.
- Talvitie, A. (1997). Things planners believe in, and things they deny, *Transportation* **24**: 1–31.
- Talvitie, A. (2001). Comment of Richard Willson's paper: assessing communicative rationality as a transportation planning paradigm, *Transportation* **28**: 207–210.

- Tillema, F. (2004). *Development of a Data Driven Land Use Transport Interaction Model*, PhD thesis, University of Twente, The Netherlands.
- Tolley, R. S. and Turton, B. J. (1995). *Transport Systems, Policy and Planning: A Geographical Approach*, Longman, Harlow, England.
- TRB (1997). Toward a sustainable future: addressing the long-term effects of motor vehicle transportation on climate and ecology, *Special report 251*, Transportation Research Board - National Research Council, Washington D.C., USA.
- UNCED (1992). Agenda 21, *Technical report*, United Nations Conference on Environment & Development (UNCED), Rio de Janeiro, Brazil.
- Van den Bergh, J. C. J. M. (1991). *Dynamic Models for Sustainable Development*, PhD thesis, Free University of Amsterdam, Amsterdam, The Netherlands.
- Van den Boom, T. and De Schutter, B. (2004). Optimization in systems and control, *Lecture Note SC4090*, Delft University of Technology, Delft, The Netherlands.
- Van Dieren, W. (ed.) (1995). *Taking Nature into Account: A Report to The Club of Rome*, Springer Verlag, New York, USA.
- Van Gent, A. L. (2005). *Aggregating Transport Networks*, Master's thesis, University of Twente, Enschede, The Netherlands.
- Van Nes, A. E. (2004). *Environment Capacity in Dutch Urban Passenger Transport*, Master's thesis, University of Twente, Enschede, The Netherlands.
- Van Wee, B. and Geurs, K. (2000). Backcasting as a tool to develop a sustainable transport scenario assuming emission reductions of 80-90%, *Innovation* **13**(1): 47-62.
- Vasconcellos, E. A. (2001). *Urban Transport, Environment and Equity: The Case for Developing Countries*, Earthscan, London, England.
- Vickerman, R. (2001). The concept of optimal transport systems, in K. J. Button and D. A. Hensher (eds), *Handbook of Transport Systems and Traffic Control*, Vol. 3 of *Handbooks in Transport*, Pergamon, Amsterdam, The Netherlands, chapter 5, pp. 47-59.
- Von Weizsäcker, E., Lovins, A. B. and Lovins, L. H. (1998). *Factor Four: Doubling Wealth, Halving Resource Use*, Earthscan, London, England.
- Wachs, M. and Kumagai, T. G. (1973). Physical accessibility as a social indicator, *Socio-Economic Planning Science* **6**: 357-379.
- Wackernagel, M. and Rees, W. (1995). *Our Ecological Footprint: Reducing Human Impact on the Earth*, New Society Publishers, Philadelphia, USA.

- WBCSD (2001). *Mobility 2001*, World Business Council for Sustainable Development, Geneva, Switzerland.
- WCED (1987). *Our Common Future*, World Commission on Environment and Development (WCED), Oxford, England.
- White, N. P. and Senior, M. L. (1983). *Transport Geography*, Longman, Harlow, England.
- Whitelegg, J. (1993). *Transport for a Sustainable Future: The Case for Europe*, John Wiley & Sons, Chichester, England.
- Whitelegg, J. (2003). Car wars, *The Guardian*, **March 26**.
- Williams, H. C. W. L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit, *Environment and Planning A* **9**: 285–344.
- Willson, R. (2001). Assessing communicative rationality as a transport planning paradigm, *Transportation* **28**: 1–31.
- World Bank (1996). *Sustainable Transport: Priorities for Policy Reform*, The World Bank, Washington D.C., USA.
- World Bank (2002). *Cities on the Move: A World Bank Urban Transport Strategy Review*, The World Bank, Washington D.C., USA.
- Wymer, C. R. (1992). Estimation of nonlinear continuous-time models from discrete data, in P. C. B. Phillips and V. B. Hall (eds), *Models, Methods and Applications in Econometrics*, Basil Blackwell, Oxford, England, pp. 91–114.
- Wymer, C. R. (2001). Wysea: Systems Estimation and Analysis Programs, *Software program*, Dr. C.R. Wymer (wymer@mail.com).
- Yang, H. (1995). Heuristic algorithms for the bilevel origin-destination matrix estimation problem, *Transportation Research* **29B**(4): 231 – 242.
- Yang, H. and Bell, M. G. H. (1997). Traffic restraint, road pricing and network equilibrium, *Transportation Research* **31B**(4): 303–314.
- Yang, H. and Bell, M. G. H. (1998). Models and algorithms for road network design: a review and some new developments, *Transport Reviews* **18**(3): 257 – 278.
- Yang, H., Bell, M. G. H. and Meng, Q. (2000). Modeling the capacity and level of service of urban transportation networks, *Transportation Research* **34B**: 255 – 275.
- Yang, H. and Huang, H. J. (1998). Principle of marginal-cost pricing: how does it work in a general network?, *Transportation Research* **32A**(1): 45–54.

- Zietsman, J. (2000). *Incorporating Sustainability Performance Measures into the Transportation Planning Process*, PhD thesis, Texas A&M University, USA.
- Zuidgeest, M. H. P., Akinyemi, E. O. and Van Maarseveen, M. F. A. M. (2001). Challenging sustainable accessibility: accessibility and mobility concepts in dynamic urban environments, *CD-ROM 9th World Conference on Transport Research (WCTR)*, Seoul, Korea.
- Zuidgeest, M. H. P. and Van Maarseveen, M. F. A. M. (2000). Transportation planning for sustainable development, *CD-ROM South-African Transport Conference - Action in Transport for the new Millennium*, CSIR Conference Centre, Pretoria, South Africa.
- Zuidgeest, M. H. P., Witbreuk, M. J. G. and Van Maarseveen, M. F. A. M. (2000). Sustainable transport: a review from a pragmatic perspective, *CD-ROM South-African Transport Conference - Action in Transport for the new Millennium*, CSIR Conference Centre, Pretoria, South Africa.

Nomenclature

General

*	indicates optimal value or composite value
\mathcal{B}	behaviour
C	controller
ε	tolerance criterium
$f(\cdot), g(\cdot)$	general functions
$F(\cdot), H(\cdot)$	general functions
\mathcal{G}	transport network
\mathcal{H}	Hamiltonian equation
J	objective function, cost criterion
\mathcal{J}	Jacobi matrix
\mathcal{L}	Lagrangean equation
L	(log-)likelihood function
$\mu(t)$	costate variable
$\eta(t)$	multiplier variable
$p_{j i}$	conditional chance on j given i
$r(t)$	residual
R^2, \tilde{R}^2	standard and corrected coefficient of determination
S	least squares criterion
$S(t)$	polynomial
\mathcal{S}	dynamic system
\mathcal{T}	time
$u(t)$	control variable
\mathcal{W}	signal space
$\omega(t)$	multiplier variable
$x(t)$	state variable

Sets

\mathcal{C}	general choice-set
\mathcal{I}	set of origins i
\mathcal{J}	set of destinations j

\mathcal{J}^i	set of destinations j , excluding destination i
\mathcal{K}	set of population segments k
\mathcal{L}	set of links l
\mathcal{M}	set of modes m
\mathcal{N}	set of nodes n
\mathcal{P}	set of pollutants p
\mathcal{R}_{ij}	set of routes r between i and j
\mathcal{R}_{ij}^l	set of routes r between i and j containing l
r	set of links l making-up route r
\mathcal{U}	set of controls u
\mathcal{Z}	set of centroids z

Roman

a_i^0	parameter intercept trip production model	[pers h ⁻¹]
a_i	parameter general gravity model	[-]
a_{ij}	relative accessibility	[pers h ⁻¹]
A, A_i	integral accessibility	[pers h ⁻¹]
A_{ij}, A_{ijm}	integral accessibility	[pers h ⁻¹]
a_1	parameter demand elasticity	[-]
b_j	parameter general gravity model	[-]
b_1	parameter demand elasticity	[-]
B	sound level traffic mix factor	[km h ⁻¹]
c_1^c	unit control cost (construction)	[€ h km ⁻¹ pcu ⁻¹]
c_2^m	unit control cost (maintenance)	[€ h km ⁻¹ pcu ⁻¹ cm ⁻¹]
c^U	unit control cost (general)	var ¹
c, c_{ij}	generalised cost of travelling	[h]
c_i^*, c_{ij}^*	composite cost of travelling	[h]
c_{ijm}^*	composite cost of travelling	[h]
\tilde{c}_{ij}^*	composite cost of travelling before measures are implemented	[h]
C_l, C_l^e	general and effective link capacity	[pcu h ⁻¹]
C_l^d	design link capacity	[pcu h ⁻¹]
C_r^P, C^P	route and overall productive capacity	[pers km h ⁻¹]
d_l, d_{ij}	link length, origin-destination distance	[km]
d_r	route distance	[km]
D^E	equivalent lane distance	[km]
D_{ij}	demand elasticity	[-]
E^p	emission of pollutant p	[kg]
E, \tilde{E}	general emission, integral emission	[kg]

¹General parameter/variable that varies.

E^*, \tilde{E}^*	point and integral environmental capacity	[kg]
\mathcal{E}^Q	equity measure	[-]
f	balancing factor	[-]
F^*, \tilde{F}^*	point and integral financial capacity	[€]
G_{ijmr}	simultaneous destination, mode, route-choice model	[-]
h, h_m^*	emission factor and composite emission factor	[g km ⁻¹]
H_i	number of households	[hh]
I_l, I_p	actual and perfect maintenance level	[m km ⁻¹]
K	kilometres driven in the network	[km]
L_l^{eq}	link equivalent sound level	[dB]
M_l	link maintenance	[-]
o_m	vehicle operating cost	[€ km ⁻¹]
$P_r^{\Delta t}$	traveller/person throughput during Δt	[pers km h ⁻¹]
P	traveller/person throughput	[pers km h ⁻²]
q_{mr}^p	person flow	[pers h ⁻¹]
q_{mr}^v, q_{lm}^v	vehicle flow	[pcu h ⁻¹]
Q, Q_i	production potential	[pers h ⁻¹]
\tilde{Q}_i, \tilde{Q}_i	corrected and elastic production potential	[pers h ⁻¹]
$\check{Q}_{i k}$	production potential share for segment k	[pers h ⁻¹]
$R_{s_{imp}}^*$	composite emission rate	[kg]
R_l	road roughness	[m km ⁻¹]
s_l	link vehicle speed	[km h ⁻¹]
S_l	modified structural number	[-]
T	general demand for travel	[pers h ⁻¹]
T^t	total number of trips generated	[pers h ⁻¹]
T_l^D	design traffic	[axles]
u_j, u_{ijmr}	utility values	[h]
U_l^c, U_l^t	capacity controls	[pcu h ⁻¹]
U_l^m	maintenance control	[cm]
U^v	vehicle tax control	[€ km ⁻¹]
U_j^p	parking tax control	[€]
v	climate coefficient	[-]
v_j, v_{ij}	observed utility values	[h]
v_{ijm}, v_{ijmr}	observed utility values	[h]
V_l	link volume	[pcu h ⁻¹]
w	road type coefficient	[-]
W_l	wear	[axles year ⁻¹]
x_n	social-economic variable	var
X_j	attraction potential	[pers h ⁻¹]

\tilde{X}_j	corrected attraction potential	[pers h ⁻¹]
z_n, z_g	zonal and exogenous characteristics	var
Greek		
α	general parameter generalised cost function	[-]
$\alpha_{k_{1i}}$	parameter trip production function	[-]
$\alpha_{k_{2j}}$	parameter trip attraction function	[-]
α_1	parameter travel time function	[-]
α_2	time-value of money parameter	[h € ⁻¹]
α_3	conversion parameter IRI to $C_l^e(t)$	[km m ⁻¹]
α_4	conversion peak hour to an equivalent day	[h]
α_5	conversion parameter $U_l^m(t)$ effectiveness to wear $M_l(t)$	[cm ⁻¹]
α_6	adjustment coefficient	[T ⁻¹]
α_7	parameter smoothing function J_{smooth}	[-]
β	general parameter generalised cost function	[h € ⁻¹]
β_1	parameter travel time function	[-]
β_{2m}	weight for mode m in accessibility function	[-]
β_{3m}	weight travel time in utility function	[-]
β_{4m}	weight travel cost in utility function	[h € ⁻¹]
β_5	pavement condition elasticity to $C_l^e(t)$	[-]
β_6	factor structural component of roughness	[-]
β_7	pavement overlay effectivity	[-]
$\beta_{8m p}$	velocity-elasticity of emission for mode m and pollution type p	[-]
β_9	parameter in cost-criterion (C2)	[-]
γ_1	adjustment coefficient	[T ⁻¹]
γ_2	adjustment coefficient	[T ⁻¹]
δ_l	level-of-service (LOS)	[-]
Δt	time period	[h]
ΔS	consumer surplus ratio	[-]
$\epsilon_{m p}$	mode m and pollutant p specific emission factor	[g km ⁻¹]
$\epsilon_j, \epsilon_{ij}$	random part utility values	[h]
$\epsilon_{ijm}, \epsilon_{ijmr}$	random part utility values	[h]
ζ	parameters regression function	[-]
θ_{1m}	reciprocal vehicle-occupancy factor	[pcu ⁻¹]
θ_{2m}	vehicle-tax conversion	[-]
κ_{ijmr}	travel cost	[€]
λ_1	scale parameter logit model	[h ⁻¹]

λ_2	scale factor composite cost to induced demand	$[\text{h}^{-1}]$
λ_3	scale factor composite cost function	$[\text{h}^{-1}]$
μ	general gravity constant	$[\text{pcuh}^{-1}]$
ν	land dampening coefficient	$[-]$
ξ_1	pavement deterioration rate	$[-]$
π_k	income-level for segmentation k	$[\text{€}]$
ρ_1	axle load equivalence exponent	$[-]$
τ_r, τ_l	route and link travel time	$[\text{h}]$
ϕ	equivalent subtending angle	$[\text{°}]$
Φ_{1m}	equivalent standard axel load factor	$[-]$
$\varphi_{my}, \varphi_y^S$	average and standard single axle y load	$[\text{tonnes}]$
χ_1	parameter penalty function J_{pen}	$[-]$
ψ	equity threshold value	$[-]$

About the author

Mark Zuidgeest was born in Den Helder, The Netherlands, on 23 January 1974. After attending 'Farel College' in Ridderkerk from 1986 to 1992, where he received his VWO (pre-university education) diploma, he enrolled the four-year Masters course in Civil Engineering & Management at the University of Twente (UT) in Enschede, The Netherlands. During his studies Mark specialised in traffic engineering and transport planning. As a traineeship, he worked for some time at TNO-INRO in Delft in 1995. In the same year he also attended a summer camp in Chimaltenango, Guatemala, where he got interested in international development. After finishing his Masters thesis in 1997 on 'hydrodynamic modelling of macroscopic traffic flow models' in the Integrated Modelling group and Centre for Transport Studies (CTS), he joined the International Institute for Infrastructure, Hydraulics and Environment (IHE-Delft) in Delft as a lecturer in transport engineering.

In IHE-Delft Mark enjoyed teaching travel demand modelling and traffic management, as well as acting as a study tour-guide through Europe to foreign students. Besides, as part of the World Bank Sub-Saharan African Transport Program (SSATP), he got to visit and work in the East-African countries Kenya and Tanzania.

In 1999 Mark returned to CTS to work on a Ph.D. thesis regarding transport models for sustainable urban development under supervision of professor Martin van Maarseveen (CTS) and professor Edward Akinyemi (University of Ilorin, Nigeria, also UNESCO-IHE-Delft), while also teaching traffic engineering and transport planning courses to, as well as supervising (international) research work of students from UT, UNESCO-IHE-Delft and the International Institute for Geo-Information Science and Earth Observation (ITC) in Enschede. In 2003 he became an assistant professor in transport engineering in CTS. His love for teaching has been awarded twice with the 'ConceptT - student society for civil engineering - teacher of the year' award for 2002 and 2003.

In 2004 he spent some time in the National Center for Transportation Studies (NCTS) in the University of the Philippines in Quezon City, The Philippines.

In his free time Mark plays underwaterhockey and dreams about living in Africa.

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: Evaluation 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

Muילerman, 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

Van Nes, R., *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 Optimisation*, 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
- Weerd, 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
- Hegy, 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 Modeling: 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 Optimisation Approach*, T2005/3, April 2005, TRAIL Thesis Series, The Netherlands