Spatial analysis of bicycle use and accident risks for cyclists

Grégory Vandenbulcke-Plasschaert

Thèse présentée en vue de l'obtention du grade de Docteur en Sciences

Université catholique de Louvain

Louvain-la-Neuve, Novembre 2011

Jury composition

Isabelle Thomas (advisor)
Professor & FRS-FNRS Research director
CORE & Department of Geography, Université catholique de Louvain

Marie-Laurence De Keersmaecker (chairman)
Professor
Department of Geography, Université catholique de Louvain

Dominique Peeters
Professor
CORE & Department of Geography, Université catholique de Louvain

Claire Dujardin
FRS-FNRS Postdoctoral researcher
CORE & Department of Geography, Université catholique de Louvain

Luc Int Panis
Professor & Research coordinator
VITO & Transportation Research Institute, Universiteit Hasselt

Ann Verhetsel
Professor
Department of Transport & Regional Economics, Universiteit Antwerp

Roger W. Vickerman
Professor
School of Economics, University of Kent

Acknowledgments

This doctoral thesis was funded by the Belgian Science Policy (Belspo), within the framework of the SHAPES research project (March 2007 – May 2011). I am grateful to Belspo for this support, as well as to the members of the follow-up committee for their numerous advices.

***

Je remercie tout d'abord très chaleureusement ma promotrice, Isabelle Thomas, qui m'a permis de me mettre rapidement en selle dans le cadre de ce parcours doctoral intense mais passionnant. Je lui suis tout particulièrement reconnaissant pour son suivi constant, son intérêt marqué pour le sujet de thèse, ses nombreux encouragements, et sa très grande disponibilité pour répondre à mes questions (disponibilité qui frôle quasi le 24h/24… les NTIC aidant…). Ses relectures de mes « romans » et ses suggestions qui en découlent sont d'ailleurs indissociables de l'aboutissement de cette thèse. En seulement six années passées au sein de l'université, l'impression d'avoir mûri est vraiment réelle…

I would also like to thank other persons, who played a prominent role within the framework of my thesis… First of all, I would like to thank all the members of the dissertation committee, i.e. Luc Int Panis, Claire Dujardin, Ann Verhetsel, Marie-Laurence de Keersmaecker, and Roger W. Vickerman. I thank all of them for having accepted evaluating my work, as well as for their fruitful comments! In particular, I would like to thank Luc Int Panis (coordinator of the SHAPES project), who has expressed huge interest towards my work from the very beginning… His interest and his enthousiasm to read my thesis was always a great source of motivation for me! Many thanks to Ann Verhetsel, for having participated at (almost) all my presentations during seminars/colloquiums, for the memorable visit of the port of Antwerp, as well as for her great human contact! Je remercie aussi Claire Dujardin pour ses encouragements, mais aussi pour ses relectures (très) minutieuses et ses conseils toujours très précieux. Sur de nombreux points, elle a souvent constitué une référence pour moi au niveau professionnel… J'ai notamment fait de sa thèse un vrai livre de chevet vers la fin de mon parcours doctoral. Mes remerciements vont également droit à Dominique Peeters, cycliste récréationnel très appliqué et dont la porte m'est toujours restée ouverte au besoin…

En dehors des membres du jury de thèse, la contribution d'un très grand nombre d'autres personnes m'a également permis de gagner un temps considérable dans le cadre de mes recherches. Je pense notamment à Julie Frère, Laurent Van Malderen, Nicolas Sougnez, Cédric Taverne, Bas de Geus, Joris Aertsens, David Dabin, Isabelle Chalanton, Florian Mayneris, et Patrick Meyfroidt. Que ce soit pour la collecte de données et/ou pour le temps qu'ils ont bien voulu me consacrer à discuter de certaines idées, leur aide m'a été très précieuse et je les en remercie de tout cœur… Je pense encore aux nombreux cyclistes qui ont bien voulu remplir le questionnaire SHAPES pendant de nombreuses semaines… Merci à eux, car sans leur aide, le quatrième chapitre de cette thèse n'aurait jamais vu le jour. Merci également à toute l'équipe administrative du CORE, et plus particulièrement à Anne-Marie Pessleux et Catherine Germain, qui ont toujours été très chaleureuses et disponibles pour moi en cas de pépin.

I would also like to thank all my colleagues and friends from CORE – and more particularly Joachim, Gauthier, Nicolas, Stéphane, and Salome – for all the great moments I shared with them as well as for the huge support they gave me at the end of my thesis. Their friendship was 'treasurable' for me as they really helped me to 'disconnect' from my thesis. Un tout grand merci aussi à tous mes collègues du département de Géographie, et plus particulièrement à Cath et Elise, pour leur très bonne compagnie, pour leurs nombreux encouragements, et pour m'avoir permis de garder pied dans le département de Géographie…

Enfin, j'aimerai remercier toutes les personnes qui comptent pour moi et qui m'ont entouré, du début à la fin de la thèse, aussi bien pendant les bons et les mauvais moments de vie. Merci à tous mes amis, pour tous vos encouragements, pour les nombreux moments de joie, d'émerveillement et de complicité que vous m'avez offerts… Que ce soit au travers de voyages, grands événements, soirées de retrouvaille, soirées jeux ou sportives, vous avez indirectement contribué à l'aboutissement de cette thèse en me permettant de m'éloigner de celle-ci. Un tout grand merci également à ma famille, pour les nombreuses tranches de rire qui m'ont toujours permis de me ressourcer, et pour m'avoir énormément soutenu pendant ma dernière année de thèse. Votre aide a été telle que je ne pense pas que j'aurai pu finir la thèse en de bonnes conditions... Les quelques lignes qui vous sont consacrées ici sont donc bien peu de choses par rapport à ce que vous avez fait pour moi…

'Last but not least', mes derniers remerciements vont à Stéphanie, pour m'avoir constamment soutenu pendant ma dernière année de thèse, pour sa patience face à mes moments de découragement et nombreuses sautes d'humeur, pour sa très grande compréhension face à cet aboutissement de projet de thèse qui me tient à cœur, et encore pour tous les projets qui restent à venir…

Table of contents

Jury composition .... **i**
Acknowledgments .... **iii**
Part I: General introduction – Research objectives and background **1**
Chapter 1: Introduction **3**
1.1 Research focus and motivations 3
1.2 Background 4
1.2.1 International context: cycling as a sustainable alternative to car use 4
1.2.2 The Belgian context 7
1.3 General objective 8
1.4 Terminology 11
1.5 Methodological approaches and gaps 13
1.5.1 Overview of current research into bicycle use and cycling accidents 13
1.5.1.1 Research into bicycle use 13
1.5.1.2 Research into cycling accidents and accidents in general 15
1.5.2 Research gaps, challenges and spatial aspects 19
1.5.2.1 Lack of reliable and high-resolution data 20
1.5.2.2 Underreporting of cycling accidents 22
1.5.2.3 Spatial data and attendant effects 23
1.5.2.4 Network phenomena and planar assumption 26
1.5.2.5 Estimation of the accident risk at the network level 27
1.6 Outline of the thesis 28
1.6.1 Outline of part II – Areal data analyses and cycle commuting (Belgium) 30
1.6.2 Outline of part III – Point data analyses along a network space and cycling accidents (Brussels) 31
1.6.3 Outline of part IV – Conclusions and policy recommendations 33
Part II: Spatial analysis of commuter cycling **35**
Chapter 2: Bicycle use and the accident risk for commuters **37**
2.1 Introduction 38
2.2 Data sources and studied area 39

2.2.1 Studied area ... 39
2.2.2 Data ... 41
2.2.2.1 Population census ... 41
2.2.2.2 Road accident statistics ... 42
2.2.2.3 Urban hierarchy ... 46
2.3 Bicycle use *versus* urban hierarchy ... 46
2.3.1 Background ... 46
2.3.2 Exploratory data analysis ... 48
2.3.2.1 Cycling, urban hierarchy and distances ... 48
2.3.2.2 Odds ratios ... 51
2.3.2.3 Exploring inter-municipality differences ... 52
2.4 Bicycle use and risk ... 54
2.4.1 Clustering municipalities ... 54
2.4.2 Analysis of the results ... 55
2.4.3 Regional and inter-municipality differences ... 57
2.5 Conclusion ... 59
Chapter 3: Spatial determinants of cycle commuting 63
3.1 Introduction ... 64
3.2 Identifying the main determinants of bicycle use ... 65
3.2.1 Demographic and socio-economic determinants ... 65
3.2.2 Cultural and societal determinants ... 66
3.2.3 Environmental determinants ... 67
3.2.4 Policy-related determinants ... 68
3.3 Objectives and data ... 71
3.3.1 Demographic and socio-economic variables ... 71
3.3.2 Environmental and policy-related factors ... 73
3.3.3 Limitations ... 75
3.4 Methodology ... 75
3.4.1 Ordinary least squares model ... 76
3.4.2 Spatial autoregressive modelling ... 76
3.4.3 Diagnostics for spatial autocorrelation ... 77
3.4.4 Diagnostics for heteroskedasticity in presence of spatial autocorrelation ... 78
3.4.5 Spatial heterogeneity and regimes ... 79
3.4.6 Diagnostics for structural instability ... 80
3.5 Results and discussion ... 81
3.5.1 Basic statistics and bivariate correlations ... 83
3.5.2 OLS results ... 83
3.5.3 Choice of the spatial weight matrix ... 86
3.5.4 Spatial lag results ... 86
3.5.5 Accounting for spatial heterogeneity ... 89

3.5.5.1 Diagnostics: Chow tests and exploratory spatial data analyses 89
3.5.5.2 Spatial regime regression with a spatially lagged variable 90
3.5.5.3 Regional variation and the relative importance of the variables 91
3.5.5.4 Spatially lagged variable 96
3.5.5.5 Analysis of the residuals 97
3.5.6 Controlling for short commuting distances and spatial interactions 99
3.5.6.1 Effect of short commuting trips (≤ 10 km) 99
3.5.6.2 Effect of spatial interactions 99
3.6 Conclusion 100
Part III: Spatial analysis of accident risks for cyclists (Brussels-Capital Region) 103
Chapter 4: Reported versus unreported cycling accidents 105
4.1 Introduction 106
4.2 Spatial context: the case of the Brussels-Capital Region 108
4.2.1 Diagnosis: mobility and cycling levels in Brussels 108
4.2.2 Why Brussels? 109
4.2.3 Spatial subareas 110
4.3 Data collection 111
4.3.1 Construction of the 'bikeable' network 111
4.3.2 Accident geocoding 112
4.3.2.1 Reported cycling accidents (DGSEI data) and the under-registration issue 112
4.3.2.2 'Unreported' cycling accidents (SHAPES survey) .. 112
4.3.2.3 Accident geocoding process 113
4.3.3 Infrastructure factors 114
4.3.3.1 Bridges and tunnels 115
4.3.3.2 Traffic-calming areas 115
4.3.3.3 Intersections (crossroads) 116
4.3.3.4 Tram tracks and public transport stops 117
4.3.3.5 Cycle facilities and discontinuities in the bicycle network 118
4.3.3.6 Parking facilities (motorised vehicles) 120
4.3.3.7 Contraflow cycling 121
4.3.3.8 Urban facilities and public services 122
4.3.4 Data limitations 122
4.4 Methodology 124

4.4.1 Comparative statistics and odds ratios 124
4.4.2 Point pattern analyses in traffic-accident research 125
4.4.2.1 Initial point pattern exploration 126
4.4.2.2 Univariate and bivariate K-function analyses 128
4.5 Results and discussion 133
4.5.1 Comparative statistics and odds ratios 133
4.5.2 Initial point pattern exploration 135
4.5.3 Network K-functions and cross K-functions 141
4.6 Conclusions 148
Chapter 5: Accident risk when cycling in Brussels 151
5.1 Introduction 152
5.2 Conceptual and methodological framework 154
5.2.1 Pre-requisites 155
5.2.2 From ecology and epidemiology… 156
5.2.2.1 Presence-only data 156
5.2.2.2 Case-control studies 157
5.2.3 … to traffic accident research 158
5.2.3.1 Exposure variable 159
5.2.3.2 Selection of controls 160
5.2.4 Modelling strategy 161
5.2.4.1 Bayes rule 161
5.2.4.2 Bayesian hierarchical modelling and accident risk model 161
5.2.4.3 Autoregressive and autologistic risk models 163
5.2.4.4 Initial values and model selection 166
5.2.4.5 Convergence diagnostics (Bayesian framework) 169
5.3 Data collection 171
5.3.1 Accident data 173
5.3.1.1 Accident geocoding ($y_i = 1$) 173
5.3.1.2 PBTI and generation of controls ($y_i = 0$) 173
5.3.2 Risk factors 177
5.3.2.1 Infrastructure factors 178
5.3.2.2 Traffic conditions 182
5.3.2.3 Environmental risk factors 184
5.3.2.4 Interaction variables and intersect analyses 185
5.3.2.5 Ignored risk factors 185
5.4 Results 186
5.4.1 Bivariate associations 186
5.4.2 Model diagnostics and selection 187
5.4.2.1 Accident risk modelling 187

5.4.2.2 Autologistic and autoregressive accident risk modelling ... 188
5.4.3 Discussion of the results of the autologistic model ... 188
5.4.3.1 Infrastructure-related risk factors ... 192
5.4.3.2 Traffic conditions ... 195
5.4.4 Predictions of the risk for a specific road trajectory: a tool for planners? ... 196
5.5 Conclusion ... 200
Part IV: General conclusions and policy implications **205**
Chapter 6: Conclusion **207**
6.1 Main findings ... 209
6.1.1 Methodological conclusions ... 209
6.1.2 Empirical conclusions ... 215
6.2 Policy implications and recommendations ... 222
6.2.1 Engineering ... 223
6.2.2 Education ... 227
6.2.3 Enforcement ... 227
6.2.4 Encouragement ... 228
6.2.5 Evaluation ... 228
6.2.6 Concluding remarks ... 229
6.3 Limitations of this thesis ... 235
6.3.1 Data limitations ... 235
6.3.2 Methodological and technical issues ... 236
6.4 Perspectives for future research ... 239
6.5 Concluding words ... 244
Appendix A: Notes to Chapter 2 **247**
Appendix A.1: The urban hierarchy of Belgian municipalities ... 247
Appendix B: Notes to Chapter 3 **249**
Appendix B.1: Variables used: description, units of measurement and data sources ... 249
Appendix B.2: Regression coefficients for the spatial regime specification ... 252
Appendix B.3: Impact of spatial interactions ... 253
Appendix C: Notes to Chapter 4 **257**
Appendix C.1: Infrastructure factors – Description and data sources ... 257
Appendix C.2: Blackspots of cycling accidents in the Pentagon ... 260
Appendix D: Notes to Chapter 5 **261**
Appendix D.1: List of risk factors ... 261
Appendix D.2: Descriptive statistics of the selected risk factors ... 267
Appendix D.3: Descriptive statistics for the continuous risk factors ... 271

Appendix D.4: Logistic model – Results from the frequentist framework 274
Appendix D.5: Logistic model – Model fit and evaluation, diagnostics and inferential tests....................275
Appendix D.6: Convergence diagnostics for the autologistic model 276
Appendix E: Publications and personal contribution 279
Appendix E.1: Approximate % of the time budget devoted to each task and chapter 280
Bibliography....................283
Data sources and on-line resources....................313
List of figures 315
List of tables....................317

Part I: General introduction – Research objectives and background

Chapter 1

Introduction

1.1 Research focus and motivations

Falling within the framework of a larger research project (SHAPES) dealing with the estimation of costs and benefits related to commuter cycling, this thesis aims at identifying some of the factors that influence the spatial variation of cycle commuting to work and cycling accidents. Our motivation is twofold. Of interest is, first, the fact that bicycle use might act as a catalyst for policies oriented towards a sustainable development of the society. It indeed holds the potential to mitigate some of the main car-related concerns with which our society is faced nowadays. Lying at the heart of sustainability-related issues, this thesis then aspires to deliver a sound scientific support for policies aiming at encouraging bicycle use and making it safer. Second, focusing on bicycle use and cycling accidents implies taking up many methodological challenges. Starting from a broad-minded standpoint, it is thus decided to position this thesis at the crossroad of the research carried out in several scientific fields, with the intent to take advantage of their respective methodological strengths to deliver robust results and policy recommendations. Quantitative and transport geography, spatial econometrics, ecology and epidemiology are some of these fields into which special attention is devoted in this thesis, because of their close connections with the spatial dimension of the data. A multidisciplinary approach, with a particular focus on space, is then opted to achieve our main goals.

The present chapter is structured as follows. It first addresses the international background to bicycle use from a societal point of view (Section 1.2.1), after which it focuses on the Belgian context since our empirical analyses are conducted in Belgium (Section 1.2.2). Section 1.3 then describes and motivates the general objectives of this thesis. In Section 1.4, some generic terms used throughout the thesis are explicitly defined. Section 1.5 briefly reviews the current literature on bicycle use and cycling accidents, and then describes some of the main challenges we decided to take up. Section 1.6 concludes this chapter by presenting the general outline of this thesis.

1.2 Background

1.2.1 International context: cycling as a sustainable alternative to car use

Most developed countries nowadays face environmental and mobility problems as a consequence of widespread car use. Partly due to long-term trends such as the increase in per capita income, car ownership has increased substantially since 1950 (Pooley and Turnbull, 2000; Rietveld, 2001). This has induced many changes, and made our societies more car dependent, leading to the progressive development of new low-density residential estates as well as commercial and industrial activities in peripheral locations (peri-urbanisation). Individuals now have higher levels of mobility and they travel more often, over larger distances, and carry out more complex trips (i.e. they undertake several activities in one trip) (Jensen, 1999; Knowles, 2006). Such a car-oriented lifestyle has however various negative impacts upon society and the environment. Among other impacts, it causes increasing congestion, air and noise pollution, vibrations, health problems (e.g. due to a lack of physical activity or the inhalation of polluting agents), space and energy consumption, traffic accidents, infrastructure costs, and accessibility problems for low-income groups (Dobruszkes and Marissal, 1994; Peirson et al., 1998; EC, 2000; Kingham et al., 2001; Bergström and Magnusson, 2003; Witlox and Tindemans, 2004; Knowles, 2006; EEA, 2007). From an economic point of view, such a popularity of car use also results in a market failure since most of these indirect external costs are borne by the society, instead of being imposed on car users (Woodcock et al., 2007). Since the last few decades, the desire to reduce the massive car use is hence growing quickly as a result of these costs and negative externalities. Although the car is still widely used for transport in our travel-demanding society, current policies are now being reappraised in favour of more sustainable modes of transport and measures are gradually taken to put a stop to the growth in car use and urban sprawl. Both in Europe as well as in an increasing number of North American towns (Larsen and El-Geneidy, 2010; Pucher et al., 2011), planners and policy makers nowadays concentrate ever-increasing efforts and attention to promote bicycle use as an effective way of reducing car dependence and its attendant negative externalities.

Such a growing interest in cycling – and, more generally, in active transport – results from the fact that it can help to achieve a variety of health, transport and environment policies oriented towards a sustainable development. It offers numerous benefits for the entire society as well as for the user itself since it is a ‘green’ and healthy alternative to commuting by car (Chapman, 2007; Woodcock et al., 2007). Cycling is indeed a cheap way of being physically active and preventing the health risks of a sedentary lifestyle, which are the second major cause of premature death in industrial countries after tobacco smoking (BMA, 1992; Pucher et al., 1999; WHO, 2002a, 2002b; Pucher and Dijkstra, 2003; de Geus, 2007, 2008a, 2008b, 2009; Buehler et al., 2011; Oja et al., 2011). When performed on a regular basis (i.e. at least 30 minutes of moderate physical activity per day), bicycle use may provide a 50% reduction in the risk of developing physical disorders related to a sedentary lifestyle (e.g. coronary heart disease, obesity and type 2 diabetes), as well as it may reduce hypertension (–30% in the risk), and psychological consequences related to inactivity (such as stress, anxiety or depression). Growing evidence from the literature also indicates that health benefits of cycling are likely to exceed (health) risks associated with its activity (i.e. the traffic injuries and the adverse health effects due to the exposure to traffic exhaust) (ERSO, 2006; Woodcock et al., 2009; de Hartog et al., 2010; Aertsens et al., 2010; Int Panis et al., 2010; Rojas-Rueda et al., 2011; Rabl and de Nazelle, 2012).

Beyond bringing direct health benefits to the cyclists, the use of the bicycle provides *indirect* health benefits for the entire society as well as an environmentally friendly alternative to the car given that it does not emit air pollutants and does not have any noise pollution impact (Pucher et al., 1999; WHO, 2002a, 2002b; Rietveld and Daniel, 2004; Gatersleben and Appleton, 2007). As suggested by EC (2000) and ERSO (2006), a substantial shift from car to alternative modes of transport – such as the bicycle – could strongly reduce the environmental and health hazards caused by air and noise pollution. For instance, it is estimated that a one-third reduction in the number of car trips from 44 to 30% in Graz (Austria) would involve a 25% reduction in pollution from motorised vehicles (EC, 2000; ERSO, 2006). Increasing evidence in the literature also indicates that air and noise pollution (caused by motorized road transport) is at the root of major health hazards, such as allergic illnesses, deficits in lung-function development in children, non-allergic respiratory diseases, or increased cardiovascular risks (WHO, 2002a, 2002b; Gauderman et al., 2007; Woodcock et al., 2007). Mitigating such forms of pollution – e.g. through a modal shift from car to bicycle – is then expected to make the environment healthier as a whole, which is even truer in urban areas and during peak hours.

Although a relation of cause and effect is not yet well-established in the literature, evidence is also growing about the fact that lower fatality rates of accidents are associated with higher levels of cycling and walking (in terms of distance travelled) (Jacobsen, 2003; Pucher and Dijkstra, 2003). Interestingly, Elvik (2009) showed that a substantial shift from motorised trips to bicycle or walking is – in theory – expected to reduce the total number of accidents. Cyclists indeed impose low (injury) risks to other road users as well as they get an improved visibility and experience in the traffic, which in turn increases the demand for cycle infrastructures and encourages even more cycling (given that the perceived safety about cycling is improved). A shift from car to bicycle is then expected to mitigate to some extent the (high) economic and social costs related to traffic accidents. Unlike motorised modes of transport, cycling also has the advantage of being a space- and energy-efficient mode of transport (Pucher et al., 1999; Rietveld, 2001; Gatersleben and Appleton, 2007; Woodcock et al., 2007). It indeed preserves non-renewable natural resources from consumption, reduces the dependence of the economy upon (imported) fossil fuels, as well as it leaves land free for future investments and reduces road congestion in urban areas (and thus, indirectly, air and noise pollution) (Litman, 2004; Krizek, 2007; Woodcock et al., 2007). Of importance is also the fact that increased cycling reduces the (high) infrastructure costs caused by massive car use and the attendant urban sprawl. At best, the promotion of cycling can even help to strengthen the economic performance of specific parts of the public transport system by attracting more consumers (through e.g. bike-and-ride). Furthermore, it is also likely to cope with the current dynamic of social exclusion generated by the unequal accessibility to different modes of transport, since the costs related to its use (e.g. in terms of maintenance, fuel consumption, parking, etc.) make it affordable for a large majority of households compared to car use (Litman, 2004; Martens, 2004; Witlox and Tindemans, 2004; Martens, 2007).

Although an increase in bicycle use results in obvious benefits for the entire society and for the cyclists themselves, there are still important barriers that deter people from cycling. In particular, the risk of having an accident – as perceived by road users – is one of the most important hurdles that discourage people from cycling (McClintock and Cleary, 1996; Pucher et al., 1999; Parkin et al., 2007; Winters et al., 2011). Except in some countries or regions benefiting from the ‘safety in numbers’ effect (owing to e.g. high levels of bicycle use, strong policy support, appropriate infrastructures, etc., as it is the case in the Netherlands or Denmark), the risk for a cyclist to be involved in a road accident is high compared with motor vehicle occupants (Elvik, 2009; Reynolds et al., 2009). As vulnerable road user, the cyclist also incurs a higher risk of injury if a motorised vehicle is involved in the accident (ERSO, 2006). Besides the medical and non-medical costs (e.g. bike repair, damaged clothes, etc.), some of the main adverse consequences associated with road traffic accidents are physical pain, possible permanent disability/invalidity, psychological complications, as well as productivity and leisure time loss (Mayou and Bryant, 2003; Aertsens et al., 2010).

Finally, other barriers discouraging the use of the bicycle are large commuting distances, steep slopes, lack of proper cycle facilities, high traffic volumes, poor accessibility to urban facilities, or company-related constraints (such as the dress code or the need to carry bulk goods). Importantly, most of the observations related to these barriers are embedded on the earth's surface and often vary in intensity over space. They also distribute seldom if ever at random over this space and – in many cases – may lead to specific spatial trends, or patterns. Conducting analyses within a geographical framework then appears to be obvious if the intent is to provide further knowledge about the factors that affect cycling as well as the risks of accident linked to its practice. Ignoring such spatial aspects would otherwise carry the risk to result in wrong inferences…

1.2.2 The Belgian context

This thesis applies to Belgium, where the popularity of cycling is high on average, although far below the levels reported in the Netherlands and Denmark (Witlox and Tindemans, 2004). At the European level (EU 15), Belgium is ranked fourth, with a bicycle share of 2.4% (in traveller-kilometres/person/year), and stands out as one of the countries with the highest share of cyclists (Denmark: 5.5%; the Netherlands: 6.7%) (EU, 2003; Rietveld and Daniel, 2004). There has been a substantial decline in the use of bicycles since 1950, as the use of cars for routine trips has increased. The bicycle is now relegated to a marginal role, and is mainly used for recreational activities: indeed, in 2001, only 6.2% of commuters regularly used a bicycle as their main mode of transport (7.4% when bicycles were integrated into a multimodal chain). This compares to 68.6% of commuters who travelled by car (Verhetsel et al., 2007). However, in recent years there have been suggestions that a cycling renaissance is occurring in Belgium, as well as in a number of other European countries (Rietveld, 2001; Witlox and Tindemans, 2004).

Compared with other modes of transport, the risk of having an accident while cycling is high in Belgium: while the bicycle share is estimated at 2.4%, cyclists account for about 9.0% of the total number of traffic fatalities (EU, 2003; Rietveld and Daniel, 2004; BRSI, 2009a). Moreover, the accident risk is estimated to be four times higher than for motor vehicle occupants (and twofold when highways are not taken into account). As shown in Figure 1.1, the risk of being killed for a cyclist is also relatively high compared with European countries reporting similar levels of cycling (such as Germany or Sweden), which suggests that bicycle use could be made safer in Belgium (BRSI, 2009a).

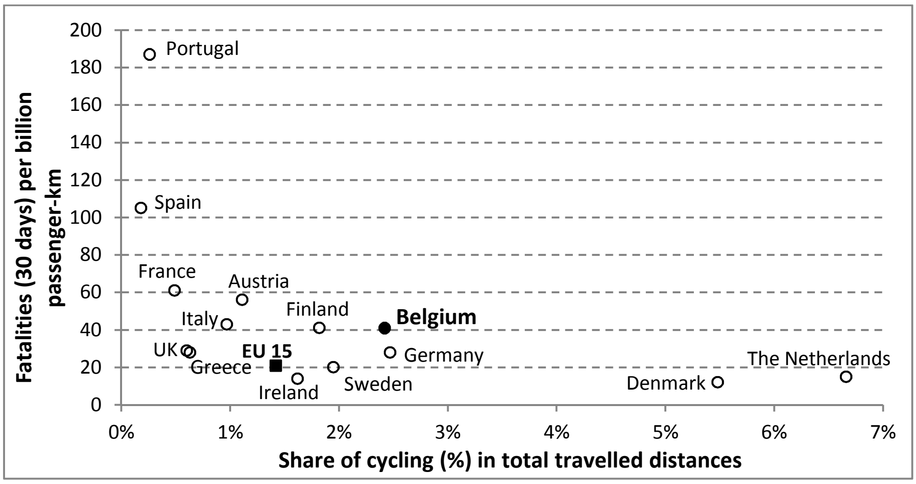


Figure 1.1: Fatality rates for cyclists and bicycle share (%) in 2000 – A European comparison. EU 15 = European Union and its 15 former member countries. Data sources: EU, 2003; Rietveld and Daniel, 2004; BRSI, 2009a.

Beyond differences between European countries (Figure 1.1), Belgium also masks strong regional differences, making it a 'fascinating' 'laboratory' for observing spatial variations in bicycle use and accident risks at a meso-scale level. As further suggested in this thesis (see Chapters 2 and 3), the northern part of the country is characterised by high levels of cycling and low risks of accidents, while the opposite situation is observed in the southern part of the country (low proportions of cyclists and high risk of accident). At an intermediate level, the Brussels-Capital Region (centrally located in Belgium) also exhibits low proportions of cyclists and high risks of accident, although fatality and serious injury risks are low for cyclists owing to the urban nature of the region. Cultural, historical, political (investments), socio-economic, demographic, and environmental factors are likely to explain to a large extent (together or separately) such strong spatial differences. As mentioned below, examining which factors significantly explain these differences constitutes one of the main challenges of this thesis.

1.3 General objective

Encouraging bicycle use requires tackling some of the main barriers to cycling through the implementation of a comprehensive package of transport and land-use policies. Barriers such as these mentioned in Section 1.2 then need to be clearly identified (and quantified) to enable policy makers and planners to develop supportive environmental conditions for more cycling. Of particular importance are also the factors that are associated with an increased/reduced probability of having a cycling accident. Identifying such factors within a scientific framework would indeed provide greater support to make bicycle use safer and, then, more common (Figure 1.2). Focussing on cycle commuting and accident risks for cyclists, the **general objective** of this thesis is then two-fold. More particularly, it aims at:

i. examining which *spatial* factors influence the spatial variation of the use of the bicycle for commuting to work at the level of the municipalities in Belgium;

ii. examining which *spatial* factors are associated with the risk of being involved in a road accident when cycling in the Brussels-Capital Region.

This thesis hence aims at contributing to the knowledge of the *spatial* determinants of cycle commuting and of one of its major deterrent factors, i.e. the risk of accident for cyclists. Obtaining further insight about such spatial determinants is of great interest as it allows identifying the main environmental/contextual factors that make a location more or less prone to the use of the bicycle (objective *i*), or more or less 'risky' for cyclists (objective *ii*). For instance, the risk of being involved in a road accident when cycling is not the same from one location to another: it spatially varies, depending on a number of inter-related factors (e.g. driver behaviour, quality of infrastructures, traffic conditions, etc.) that determine this risk and explain why some locations are more prone to generate accidents than others. Accumulated knowledge about these factors then allows establishing sound recommendations intended for policy makers and planners, especially by pinpointing locations where measures should be taken to encourage bicycle use and make it safer.

From a conceptual point of view (Figure 1.2), measures resulting from our two-fold objective may act as interrelated parts of a *virtuous circle* in which pro-cycling strategies and improvements in the bicyclist's safety may support and influence each other, thus contributing to continuously increase bicycle use and make it safer (until time *t*, after which all the 'cycling potential' is assumed to be fully exploited). More concretely, it means on the one hand that improving the safety and convenience of cycling is of prime importance to encourage the use of the bicycle as it is well-known that the (perceived) risk of cycling accident strongly deters it. Indeed, reducing the actual risk of accident – e.g. through an appropriate package of policy measures implemented at target locations – holds the potential to lower the individuals' overall perception of the risk associated with cycling, which in turn encourages even more people cycling.

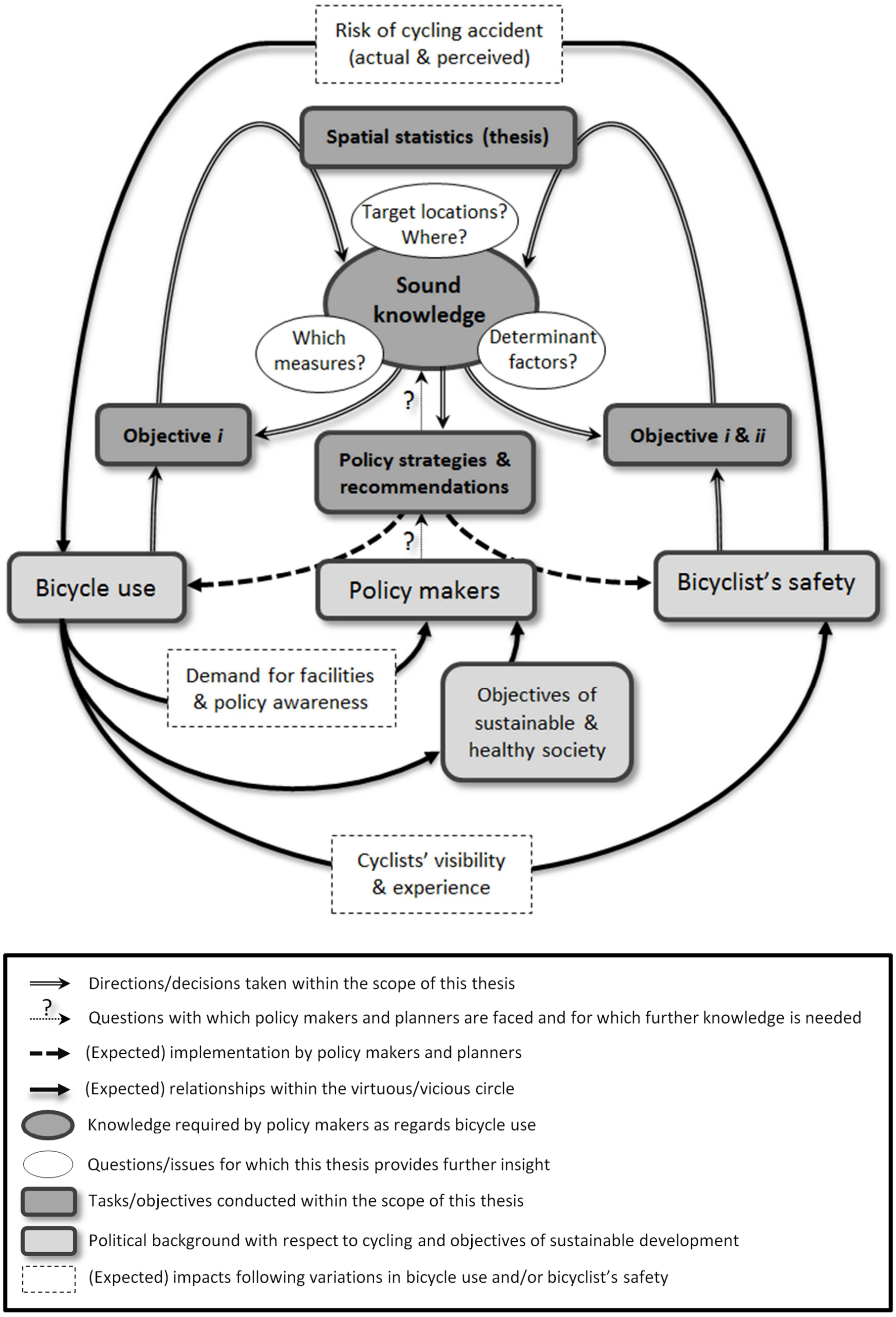


Figure 1.2: General objectives of the thesis (*i* & *ii*) within the contextual framework

On the other hand, increasing bicycle use through a comprehensive set of pro-cycling measures may also help to create a 'safety in numbers' effect (Jacobsen, 2003). The growing or the extensive use of the bicycle in a given environment may indeed be an efficient way to lead potential users questioning about cycling here (e.g. how safe and convenient is cycling here?). This may in turn encourage even more people cycling ('mass effect') and then may improve the safety of all cyclists through e.g. a greater visibility and experience of these latter in the traffic, or a better availability of high-quality cycle infrastructures (due to a higher demand and the achievement of a critical mass of cyclists). Concentrating this thesis on both interrelated aspects is then far from being unsubstantiated and aims at providing further knowledge for setting such a virtuous circle in motion (or at least maintaining it) and for achieving the objectives of sustainable development of the society.

From a methodological point of view, this thesis devotes special attention to *spatial approaches*, which are techniques centred on geographical aspects and using *spatial data/observations* with the aim to explore and/or account for their (eventual) *spatial relationships* and their attendant *spatial effects* or *biases* (see Section 1.5 for further information on these effects/biases encountered throughout this thesis). It is hence questioned throughout the general objective how do cycle commuting and accidents distribute over space, and why do they tend to be more/less frequent *in some places* than others. In other words, it is here aimed at shedding light on the (spatial) factors that explain why cycle commuting and accidents are more/less frequent in some places than others. Such spatial approaches are hence of great interest for planners and policy makers since they allow: (*i*) identifying which types of environments encourage or deter cycling (Objective *i*); and (*ii*) which factors (i.e. road infrastructures, activities, etc.) increase or lower the risk of having a cycling accident. Note that further details on these approaches are provided in Section 1.5.2.

1.4 Terminology

Before going further in describing the main methodological approaches and gaps encountered in the literature, this section aims at providing basic definitions for some of the key terms used throughout this thesis. The list of these terms is as follows:

- **Bicycle use** – also referred here to as '(bi-)cycling'– is the use of the bicycle for utilitarian (e.g. work, school) or recreational purposes (e.g. sport, racing, recreation/leisure, etc.). People making use of the bicycle for their trips are here referred as 'cyclists' or 'bicyclists'.

- **Commuter cycling** is the use of the bicycle as mode of transport for a *regular* travel between the place of residence and the workplace or school. People that use the bicycle for commuting are hence referred as 'commuter cyclists'. Note that only commuting trips to work are considered in the second part of this thesis (Chapters 2 and 3).
- A **road accident** (or traffic accident/crash) is defined as any accident occurring on a public road and involving at least one road user. In Belgium, road accidents resulting in injuries and fatalities are officially reported by the police and afterwards compiled by the Directorate-General Statistics and Economic Information (DGSEI). The DGSEI defines the severity of accidents as follows:
  - *Fatal accident*: any accident resulting in one or more road users being killed, either at the location of the accident or within the 30 days (due to the accident-related injuries);
  - *Severe/serious accident*: any accident resulting in one or more road users being seriously injured and for who an hospitalisation of more than 24 hours was reported;
  - *Slight accident*: any accident resulting in one or more road users being slightly injured and for who the hospitalisation was less than 24 hours.

  Most accident data used here come from the DGSEI. As further mentioned in this thesis, accident data generally face with a number of drawbacks that can bias the statistical inference. In particular, some of the main limitations with which this thesis is confronted are the underreporting of accidents (especially slight accidents), the encoding errors (as regards e.g. the location of the accident), and the absence of trip purpose (road users involved in the accident are not asked to register their trip purpose).
- A **casualty** is defined as any person injured or killed as a result of a slight, serious or fatal accident. Also note that a **fatality** is defined as any person killed as a result of a fatal accident.
- A **bicycle accident**, or cycling accident, is defined as any road/traffic accident occurring on a public road and involving at least one cyclist. Note that slight bicycle accidents are particularly prone to the underreporting issue, as cyclists generally do not feel the need to register their accident (because of the slight injuries, low material damages, etc.). Further information about such underreporting issue is provided in several sections of this thesis (Section 1.5.2.2, Section 2.2.2.2 and Section 4.3.2.1).
- The notion of **risk** is here defined as the probability that the outcome of interest will occur, following a particular exposure of the population or study group (Burt, 2001; Porta, 2008). In particular, the *risk of having a cycling accident* is the probability that this accident will occur, following the exposure of the cyclists in the traffic during a specified period of time (or for a specific distance).

1.5 Methodological approaches and gaps

This section aims at laying down the foundations underlying the methods used within the framework of this thesis. It first provides a short overview of research into bicycle use and cycling accidents, as well as on the statistical techniques customarily used in the scientific literature to identify the determinants of cycle commuting and accidents (Section 1.5.1). In a second step, it describes some of the main research issues we decided to address within the framework of this thesis in order to provide thorough and innovative recommendations for planners and decision makers (Section 1.5.2).

Note that it is not the goal here to provide an exhaustive review of the main theoretical concepts. Instead of falling into one methodological framework, this thesis proposes *various frameworks* suited to the individual objectives of the chapters. As a result, the purpose of this section is rather to briefly review the main methods of analysis used in the existing research into cycle commuting and accidents, after which it aims at identifying some of the main research gaps related to these methods. Further information on the methodological solutions selected here to address these gaps is provided subsequently in this thesis (notably within the framework of chapters 3, 4 and 5).

1.5.1 Overview of current research into bicycle use and cycling accidents

1.5.1.1 Research into bicycle use

Although a large number of empirical studies focus on mode choice and trip frequency, it is noteworthy that – throughout the scientific research in transport – relatively limited attention has been devoted specifically to the use of the bicycle for commuting to work, especially when compared to motorised vehicles. Overall, the bulk of studies on cycling either focuses on bicycle use *in general*, or examines the mode choice within the framework of *all* commuting trips (Heinen et al., 2010). In the former case, bicycle use is studied regardless of the trip purpose (i.e. leisure, school, work, shopping, etc.), while in the latter case the choice of commuting by bicycle is investigated considering that other modes of transport are available as alternatives to cycling (one commuter will then choose to travel on the mode which gives him/her the highest 'utility').

Much of this research into cycling is conducted throughout studies related to transportation, social sciences as well as medical and health education matters. In line with the general objective (*i*) of this thesis, most of these studies commonly aim at examining the relationship between the travel behaviour of people or commuters (measured as trip frequency, flows, mode share, mode choice, etc.) and a number of factors/determinants that are assumed to be 'explanatory' with respect to these behaviours. Invariably, the purpose is thus to identify which factors have the greatest influence on the use of a specific mode within the framework of a definite trip purpose. Based on an exhaustive literature review carried out by Heinen et al. (2010) and in Chapter 3 of this thesis, evidence from the academic research shows that travel behaviour – and, in particular, bicycle use – is influenced by factors related to the built environment (e.g. population densities, presence of cycle facilities, distances, etc.), as well as by socio-economic (e.g. income, education, car ownership, etc.), demographic (e.g. age, gender, etc.), psychological (e.g. attitudes, social norms, habits, etc.), physical (e.g. weather, hilliness, etc.), as well as safety- and cost-related factors (e.g. perceived safety, travel time, physical effort, etc.). Other studies also aim at scrutinising the preferences of cyclists with respect to specific routes (e.g. direct *vs.* safe routes) or factors (e.g. separated *vs.* on-road cycle facilities) (*ibid.*).

From a methodological point of view, empirical works traditionally use *statistical analyses* to explore and/or explain the travel behaviours and preferences of commuters (which are here referred as '*dependent*' variables) as a function of a set of explanatory factors/determinants ('*independent*' variables). Roughly, such statistical analyses can be categorised into two complementary groups. Firstly, the *exploratory analyses* generally consist of univariate or bivariate techniques that explore the data through simple graphical approaches (e.g. histograms, boxplots, scatterplots, maps, etc.) and/or using basic statistics, such as descriptive statistics (e.g. mean, standard deviation, etc.) or test statistics (e.g. Chi-Square test, Wilcoxon test, *t*-test, etc.). They also include multivariate techniques that aim at investigating the relationships/correlations existing between the different factors (using e.g. the Principal Component Analysis) or group the observations based on a set of factors (using e.g. a cluster analysis, such as the one applied in Chapter 2). Hence, such exploratory analyses mostly aim at inspecting the variables and their relationships, and sometimes precede more robust techniques (such as regression models) by highlighting the effect some factors/determinants might have on travel behaviours and preferences. Examples of exploratory analyses are legion throughout the research into bicycle use (see e.g. Dickinson et al., 2003; Witlox and Tindemans, 2004; de Geus, 2007; Dill and Voros, 2007; Gatersleben and Appleton, 2007). Secondly, a significant deal of the research is conducted within an *explanatory* framework, i.e. within which several factors/variables are considered. It generally makes use of statistical regression models, which aim at examining the travel behaviour of commuters as well as the *relative* relevance/significance of factors in explaining such travel behaviours. Different types of models are used in the literature, depending on how the dependent variable is defined (i.e. whether this latter is discrete or continuous). Logit models, probit models, linear regression models, and structural equation models are some of the most commonly used specifications (Heinen et al., 2010). In particular, binominal/multinomial logit models are based on discrete responses and are frequently used in mode choice research to evaluate the 'utility' one commuter gives to each mode of transport, as a function of a set of factors (see e.g. Noland and Kunreutheur, 1995; Wardman et al., 1997; Rodriguez and Joo, 2004). Similarly, empirical studies based on stated preference surveys also make use of such models to determine the preferences cyclists have for specific routes or facilities (see e.g. Stinson and Bhat, 2003, 2005; Hunt and Abraham, 2007). Regarding continuous dependent variables, the existing research commonly uses regression models in order to identify which determinants have a significant effect on the share of cycling or on cycle flows (see e.g. Emmerson et al., 1998; Rietveld and Daniel, 2004).

Finally, the selection of a particular model also depends on the level at which the data are available. Except for studies based on authors' own surveys, data are not always reported at the individual level and are often aggregated to areas, or zones (such as municipalities, agglomerations, towns or countries). When available in aggregate form (e.g. variation in the share of cycling per municipality), the dependent variable is continuous and may then require an appropriate model specification (such as a linear regression model in the simplest case). At the opposite, studies conducted at the individual level are generally based on discrete dependent variables (e.g. cyclist *vs.* non-cyclist) and then aim at predicting mode choice and cyclists' preferences using logistic specifications. More importantly, different results may be obtained for the same statistical analysis depending on the level at which the data are reported or aggregated. This latter issue is known as the *Modifiable Areal Unit Problem* (Openshaw, 1984) and is well-documented in the literature (see e.g. Bailey and Gatrell, 1995; Fotheringham et al., 2000). Further details on this methodological concern are provided in Section 1.5.2 of this thesis.

1.5.1.2 Research into cycling accidents and accidents in general

Cycling accidents – and road accidents in general – result from the combination and interaction between five categories of factors: human factors (e.g. driver behaviour, driver error, response to stimuli, etc.), vehicle-related factors (e.g. size or state of the vehicle), infrastructure factors (e.g. crossroad design, pavement type), traffic conditions (e.g. density, speed), and environmental factors (e.g. lighting, weather) (Miaou et al., 2003; Li et al., 2007; BRSI, 2008). Despite the fact that considerable methodological improvements have been achieved in traffic accident research during the last decades, the lack of accurate information about the human factors and accident mechanisms (e.g. acceleration, braking, etc.) as well as the driver-related privacy issues have often hampered researchers to enhance their knowledge about the exact cause and effect relationships with regard to the road accidents as a whole (Lord and Mannering, 2010). As a consequence of such data limitations, the body of the literature mainly focuses on examining the factors that affect either the *frequency* or the *severity* of accidents. Other studies aim at investigating the association between the type of *collision* (e.g. rear-end accident, side accident, etc.) and a set of factors related to the accident mechanisms (Noland and Quddus, 2004; Lord and Mannering, 2010).

For a number of reasons (e.g. privacy issues, administrative convenience, etc.), most of studies also aggregate the accidents *over space and/or over some period of time* (Aguero-Valverde and Jovanis, 2006; Liu and Jarrett, 2008; Quddus, 2008; Lord and Mannering, 2010). Regarding the aggregation over time, a period of one to several years (e.g. 3, 4 or 5 years) is generally chosen and may provide an adequate basis for further statistical analysis, although some studies may focus on shorter periods of time (several weeks or months). When accidents are aggregated over space, typical spatial units used throughout the literature are road nodes (intersections), road links (junctions) and administrative areas (such as statistical wards, municipalities, counties, regions or countries).

From a methodological point of view, much of the research into traffic accidents may be broadly classified into two groups (exploratory *vs.* explanatory models), depending on the purpose of the study. As for the research into cycling, **exploratory methods** may be used as an initial step to 'look at' the data, before performing explanatory methods. They aim at describing the accident data set using basic statistics (i.e. descriptive statistics, test statistics, odds ratios, etc.) and/or various spatial approaches. Importantly, the choice of one specific exploratory spatial method is strongly conditioned upon the level at which the data are available or aggregated (i.e. individual or spatially aggregated):

- At an *individual level*, spatial point pattern analyses are generally carried out to explore the accident data. These mainly consist of methods measuring the global variation in the mean value of the point pattern (first-order effects) or examining the tendency for local deviations from the mean value caused by the spatial correlation structure of the pattern (second-order effects) (Bailey and Gatrell, 1995; O'Sullivan and Unwin, 2002). Examples of methods exploring the first-order effects of the point pattern are centrographic techniques, (network) quadrat count analyses and (network) kernel density estimates (see e.g. Bailey and Gatrell, 1995; Levine et al., 1995a; Banos and Huguenin-Richard, 2000; Fotheringham et al., 2000; Myint, 2008; Shiode, 2008; Okabe et al., 2009). Second-order effects, on the other hand, are examined using e.g. nearest neighbour distances and (cross) *K*-function methods (see e.g. Bailey and Gatrell, 1995; Fotheringham et al., 2000). Most of these exploratory methods are presented in Chapter 4 of this thesis.

- At a *spatially-aggregated level*, segment- and area-based methods are commonly adopted to explore primarily the second-order effects in the accident data (and, to a lesser extent, first-order effects; see e.g. Lassarre and Thomas (2005)). Although they strongly depend on the definition of the neighbourhood, such methods have the advantage to pinpoint which parts of the network or which areas/zones show *statistically significant* concentrations of road accidents. In other words, they identify the significant black zones of accidents along the road network (on specific segments) or at the scale of areas/zones. On the one hand, the computation of dangerousness indices and local indicators of network-constrained spatial autocorrelation are some of the most frequently used segment-based method to detect black zones on the network (see e.g. Thomas, 1996; Black and Thomas, 1998; Flahaut et al., 2003; Steenberghen et al., 2004, 2010; Yamada and Thill, 2010). Area-based analyses, on the other hand, may be performed using Moran's *I* indices (Moran, 1948) and/or Getis-Ord Gi* local statistics (Getis and Ord, 1992) to identify black zones. *Z*-scores are then used to test the statistical significance of the computed values, and then also the statistical significance of black zones. To our knowledge, such methods are rarely applied on accident data in the literature, although some recent examples can be found in Khan et al. (2008) and Kingham et al. (2011).

Besides exploratory methods, **explanatory models** are commonly used to estimate the relative importance several factors may have on the occurrence and severity of accidents. Overall, three types of models are generally identified in the literature: the *accident-frequency models* (also referred as 'accident-count' models), the *accident-collision models*, and the *accident-severity models*. Concretely, the first category of model is generally applied to compute the probability of observing a definite number of accidents as a function of a set of accident-related factors (e.g. characteristics of the accident location, time of the accident, road users involved in the accident, etc.), while the second and third types of model overall focus on estimating the probability that an accident falls into one definite class of collision or injury severity, respectively (still as a function of a set of accident-related factors and conditional on the fact that the accident has occurred) (Ye and Lord, 2011). Among the first class of models, Poisson and Poisson-gamma (or negative binomial) models are the most common choices in the literature as accident-frequency data are Poisson-distributed and consist of non-negative integers (which precludes using models based on continuous dependent variables, such as ordinary least-square regressions) (Lord et al., 2005; Lord and Mannering, 2010). During the last two decades, other types of models have also been developed to address important data and methodological issues identified throughout the literature on accident-frequency models. Zero-inflated Poisson and negative binomial models, for instance, handle data that are characterised by a large number of zero-accident observations (or more zeros than Poisson or Poisson-gamma models would expect). Other types of models also account for various types of issues, such as under-dispersed data (Gamma models), temporal correlation (Generalised estimating equation models), or non-linear variable interactions (Generalised additive models). Last but not least, some of the accident-frequency models may even handle several issues at the same time. In particular, the Conway-Maxwell-Poisson model may address both over- and under-dispersion issues in the data, while the random-effect models turn out to be useful to treat both spatial and temporal correlations. For information purposes, readers are here urged to refer to Lord and Mannering (2010) if they are interested to get a more complete review of the literature about accident-frequency models (and their related issues). As regards the two remaining categories of models (accident-severity and accident-collision models), binomial logistic specifications are widely used throughout the literature when the dependent variable is in a binary form (e.g. fatal *vs.* non-fatal accident), while multinomial or ordered logit specifications are generally performed when multiple categories are available (e.g. no injury, slight injury, serious injury and fatal injury in the case where the responses are ordered).

Focussing on cycling accidents in particular, it turns out that much of the empirical work is recent (90's) and is mainly conducted in social sciences, medical and health care research, and transportation (including traffic accident analysis, injury prevention, transport geography and engineering) (see Eluru et al. (2008) and Reynolds et al. (2009) for a review of the literature). Examples of accident-frequency models applied to cycling accidents can be found in Wang and Nihan (2004), Hels and Orozova-Bekkevold (2007), and Schepers et al. (2011). On the other hand, empirical works aiming at comparing the impact of factors on different levels of injury severity for cyclists are far more common and can be found notably in Rodgers (1997), Klop and Khattak (1999), Kim et al. (2007) and Eluru et al. (2008). As regards accident-collision models, much of the work is – to our knowledge – quite recent and mainly aims at finding associations between the type of collision/manoeuvre (e.g. door-related accidents, rear-end accidents, overtaking accident, etc.) and a set of factors. Relevant examples can be found in Pai (2011) and Yan et al. (2011).

1.5.2 Research gaps, challenges and spatial aspects

Research into bicycle use and cycling accidents still constitutes a great challenge for transport scientists. Important data and methodological limitations have been identified as potential sources of bias, in the sense that these may lead to invalid inferences/results with respect to the explanatory factors (e.g. biased parameter estimates can be obtained). Of concern is notably the *lack of reliable data* on the factors specific to the bicycle (e.g. cycle facilities), as well as the limited attention devoted by the researchers to the presence of *spatial autocorrelation* in the data. Markedly, this latter issue is seldom if ever addressed in research into bicycle use and cycling accidents (despite the fact that many studies consider spatial information). *Underreporting of cycling accidents* is also a well-known issue in the literature, which is likely to affect the results. Given that the large bulk of studies only account for cycling accidents reported by official statistics (which constitutes the tip of the iceberg), it would be worth to question how such reported cycling accidents are representative of the unreported ones (in terms of environmental features/factors and spatially). To our knowledge, no research has been conducted yet to get such insight. Another concern is the fact that cycling accidents – but also road accidents in general – are events that are constrained to occur on a *network space*, which is not always taken into account in a number of (exploratory) studies. Last but not least, the *risk of cycling accident* associated with some definite locations or infrastructures is seldom if ever estimated in the literature on traffic accidents. Partly because there is no reliable exposure variable (e.g. bicycle flows), the strand of the literature is limited to examine the impact of several factors on accident frequency and/or on various levels of injury severity. On the other hand, surveys aiming at estimating such a risk often fail to select valid controls and raise questions about their relevance in providing consistent parameter estimates and recommendations about explanatory risk factors (see e.g. Lusk et al. (2011) and related comments).

This thesis then aims at addressing some of the aspects related to these gaps. The intent is to provide sound statistical results and, then, well-founded recommendations for policy makers and planners. Such gaps – as well as their attendant solutions – are introduced one by one in the following subsections.

1.5.2.1 Lack of reliable and high-resolution data

Most data compiled in transportation research are car-based and put little emphasis on non-motorised transports in general (i.e. walking and cycling). In particular, data on cycling are often collected in short supply and/or are generally of limited quality (Iacono et al., 2010). Heinen et al. (2010) also point out the fact that bicycle-specific factors, i.e. those *directly* influencing bicycle use and cycling accidents (e.g. cycle facilities, hilliness, etc.), are often neglected in the literature, although they would be worth considering since they are likely to provide sounder recommendations on how encouraging cycling and making it safer. Of concern is mostly the fact that such bicycle-specific factors are seldom registered at a local scale. Partly because of time and cost constraints, these are commonly aggregated over spatial units and then often impose to carry out empirical analyses over areal units, despite the fact that some other data might be available at an individual level through surveys or censuses (e.g. socio-economic and demographic data about cyclists; see Chapter 3). Such a lack of reliable and high-resolution data about bicycle-specific factors is hence one of the most central issues in research into cycling, as it often hampers to get in-depth knowledge on the factors that significantly influence both bicycle use and cycling accidents.

More importantly, conducting statistical analyses over areal units in turn requires being aware that incorrect inferences may result from the aggregation of the data, especially as regards the cycling accidents (see Section 1.5.1.2). Also, the obtained results may be conditional upon the definition of the areas/zones for which these data are spatially aggregated (Bailey and Gatrell, 1995; Fotheringham, 2000). For a same statistical analysis, different results can indeed be obtained when different levels of spatial resolution are chosen (Fotheringham, 2000). Such a sensitivity in the results has been previously demonstrated for both bivariate and multivariate analyses as well as for spatial modelling (see e.g. Openshaw and Taylor, 1979; Fotheringham and Wong, 1991; Fotheringham et al., 1995). This methodological issue is referred in the literature to as the *Modifiable Areal Unit Problem* (Openshaw, 1984).

In the light of all these issues, both Chapters 2 and 3 pay particular attention on collecting an exhaustive data set considering both ‘traditional’ factors (i.e. factors that are traditionally used in the literature, such as socio-economic factors) and bicycle-specific factors (such as cycle facilities, hilliness, motorised traffic volume, etc.). More interestingly, within the framework of Chapters 4 and 5, high-resolution data are created in order to avoid the statistical biases that could result from aggregating the data on areas or segments. Such data are mostly infrastructure-related (e.g. tram tracks, cycle facilities, etc.) and are digitised into a Geographic Information System (GIS) for the Brussels-Capital Region.

From a methodological point of view, spatial autoregressive methods (in particular, the spatial error model) may turn out to be useful in suggesting the presence of omitted data and, subsequently, in accounting for such unmeasured information by specifying a spatial autoregressive process for the error terms (see Chapter 3 for further information). The use of multilevel statistical models could also be here of interest as it could allow handling our data characterised by a hierarchical/multilevel structure, i.e. characterised by different levels of hierarchy/aggregation (Schwenkglenks, 2007; Corrado and Fingleton, 2011). For instance, observations for workplaces (1st level of the hierarchy) may be nested within municipalities (2nd level), which are in turn nested within regions (3rd level, which is here the highest level of the hierarchy) (see e.g. Vanoutrive et al., 2010 for a recent application to Belgium). Multilevel models have the advantage of separating the 'contextual effects' (i.e. the effects of group-level characteristics, or neighbourhood effects) from the 'compositional' ones (i.e. the effects of individual-level characteristics) (Duncan et al., 1998; Mohan et al., 2005; French and Jones, 2006; Johnston et al., 2007). Compared to conventional regression methods (e.g. ordinary-least squares (OLS) methods), multilevel models are also shown to result in better statistical efficiency, better identification of effects and unbiased standard errors (Goldstein, 1999; Rice, 2001). Given that individuals are grouped/aggregated on different levels of hierarchy, they also allow accounting for the dependence between observations living in a same location[1], as well as they allow modelling spatial dependence through the error term by applying e.g. the feasible generalized spatial two-stage least squares (FG2SLS) method (see e.g. Corrado and Fingleton (2011) for further information). Multilevel models however lead to a number of drawbacks and are still a subject for debate (see e.g. Oakes, 2004, 2006, 2009). Several authors indeed express concerns about the causal interpration of the effects obtained and recommend that the results of multilevel models should not be interpreted causally, especially when observational data are used (Draper, 1995; Oakes, 2004, 2009; Gelman, 2006; Gelman and Hill, 2007; Gelman et al., 2007). In such (observational) cases, Oakes (2004) argues that multilevel models do not permit to distinguish the contextual/neighbourhood effects from compositional ones. Oakes and colleagues (Oakes, 2004; Hearst et al., 2008; Johnson et al., 2008) also revealed that inferences are strongly dependent on the model assumptions, and not on the data. Among other disadvantages, multilevel models

[1] For instance, individuals coming from a same location are more likely to share the same characteristics than individuals drawn at random from the entire population. This dependence between observations violates the assumption of standard OLS regression methods as regards the independence between observations/individuals (Goldstein, 1999).

also involve complex modelling processes and have quite large sample size requirements at all levels of the hierarchy (Rice and Jones, 1997; Diez-Roux, 2000; Greenland, 2000; Hox, 2002; Schwenkglenks, 2007). Within the framework of this thesis, multilevel models are not used because of such drawbacks. Instead, it is here decided to carry out a spatial econometric framework incorporating the higher level unit(s) as dummy variable(s) in the final model. It then leads to a 'spatial regime specification', in which the higher level unit(s) is (are) determined on the basis of statistical indicators of spatial association between the observations (see Chapter 3).

1.5.2.2 Underreporting of cycling accidents

Underreporting of road accidents strongly depends on accident severity, vehicle type, age and role of the victims (i.e. passenger or driver), and the number of vehicles involved (Hauer and Hakkert, 1988; ERSO, 2006; Ye and Lord, 2011). Throughout the literature, it is well-known that cycling accidents are strongly underreported by the police, compared to motorised modes of transport. On the basis of hospital or survey data, several authors estimate that about 15% of the cycling accidents are reported by official statistics in Belgium (see e.g. Hubert and Toint, 2002; Lammar and Hens, 2004; De Mol and Lammar, 2006). Such low registration rates are explained by the fact that most cycling accidents are single-vehicle accidents and generally result in slight injuries and/or material damages only. This implies that the cyclist often does not feel the need to call the police, and then that there is no official record of the accident.

From a methodological point of view, such incomplete accident reporting may result in biased results regarding the probability of falling into one specific level of injury severity. In other words, the probability of being seriously injured in a cycling accident is overestimated, whereas this of being slightly injured is underestimated (Ye and Lord, 2011). Biased parameter estimates can also be obtained when underreporting is not taken into account in accident-frequency and accident-severity models, which implies that erroneous inferences can be made about the relative impact of the explanatory variables (Kumara and Chin, 2005; Yamamoto et al., 2008; Ma, 2009; Lord and Mannering, 2010; Ye and Lord, 2011).

As suggested by Ye and Lord (2011), the Weighted Exogenous Sample Maximum Likelihood Estimator (WESMLE) could be used as a suitable solution to improve the results when full or partial information is available about the rate of underreporting for each level of injury severity. This method indeed gives better results than methods based on Maximum Likelihood estimations and ignoring the underreporting issue (see Ye and Lord (2011) for further details).

However, it is here questioned whether or not reported and unreported accidents exhibit different characteristics/risk factors, i.e. whether or not some risk factors are specifically related to the underreporting issue itself. It matters in the sense that parameter estimates could still be biased in the case where differences (in terms of factors) are reported between reported and unreported accidents. Over- or underestimation of estimates would indeed still result from the WESMLE method if unreported accidents are associated with some specific factors (hidden/latent or not). For instance, it is not unlikely that the proportion of slight injuries (among all injuries) reported in traffic-calming areas is lower than this left unreported in such areas, which implies that the effect of traffic-calming areas on the probability of falling into a lower class of injury severity is still likely to be under-estimated with the WESMLE method (whereas it would be overestimated for the most severe classes of injuries). As a consequence, it would be hence of interest to examine whether or not reported and unreported accidents exhibit differences in terms of location and characteristics/factors (e.g. with regard to the neighbouring built environment). In this thesis, Chapter 4 focuses on this latter question by adopting a geographical point of view: it not only compares the spatial patterns of both reported and unreported cycling accidents, but it also examines if they both spatially cluster around similar spatial factors/characteristics. Throughout the evaluation of the differences between reported and unreported cycling accidents, such an analysis would in turn give a first insight in what could be the (importance of the) bias caused by the underreporting of cycling accidents on modelling results (such as these obtained in Chapter 5).

1.5.2.3 Spatial data and attendant effects

Many observations or data are inherently spatial, i.e. they possess a definite spatial reference or location on the earth’s surface. Although they are far from being common in the whole academic literature, empirical studies accounting for such a spatial dimension are in ever greater numbers in scientific disciplines where space lies at the heart of the research (such as social sciences, regional sciences and economics, epidemiology, and ecology). Compared to non-spatial data, spatial data are indeed unusual in the sense they may give rise to two well-known types of spatial effects: spatial autocorrelation and spatial heterogeneity (Anselin, 1992). Importantly, such effects violate assumptions of the classical regression models and may make biased and inconsistent the statistical inference procedure if they are ignored (Anselin, 1988; Long and Ervin, 2000). Spatial statistics then turn out to be useful in accounting for such effects. There is an abundant research into such methods throughout the literature (see e.g. Cliff and Ord, 1973, 1981; Upton and Fingleton, 1985; Griffith, 1987; Anselin, 1988;

Haining, 1990; Cressie, 1993; Bailey and Gatrell, 1995; Fotheringham et al., 2000; Lawson, 2009) but it is not the aim here to provide a complete review of the topic. We here briefly describe both spatial effects, with a particular attention on the literature in spatial econometrics.

On the one hand, *spatial autocorrelation* – also referred to as spatial dependence or spatial association[2] – follows directly from Tobler's first law of Geography (Tobler, 1979), which states that "everything is related to everything else, but near things are more related than distant things". In other words, observations will tend to exhibit similar values of factors/variables in nearby locations, leading to groups or 'spatial clusters' (characterised by spatially autocorrelated values). For instance, it is well-known that high crime areas are often surrounded by other high crime environments. Such a (positive) correlation of the values of one variable/factor over space implies that observations are not independent from each other over space. This hence violates the assumption of independently and identically distributed (i.i.d.) errors of most standard statistical models, which inflates type I errors (due to underestimated standard errors) and leads to biased and inconsistent estimates as well as poorer fit compared to the case where errors are i.i.d. (Anselin, 1992; Legendre, 1993; Legendre et al., 2002). Among other statistics, the Moran's I statistic and the Lagrange Multipliers (as well as their robust forms) can be used to detect the presence of residual spatial autocorrelation in the model. Methods that account for the presence of spatial autocorrelation when analysing spatial data are manifold and mostly include autoregressive methods (e.g. autologistic models, conditional autoregressive (CAR) models, simultaneous autoregressive (SAR) models), geostatistical methods (e.g. regression kriging, co-kriging), parameter estimation methods (e.g. generalised linear mixed models (GLMM), generalised estimation equations (GEE)) and other methods such as modified specifications of the previous models (see Miller et al. (2007) or Dormann et al. (2007) for a complete review of the literature).

On the other hand, *spatial heterogeneity* also affects the statistical validity of the model throughout the presence of non-constant variance in the errors (heteroskedasticity) and/or structural instability in the estimates between different spatial subsets of data. Accounting for both forms of spatial heterogeneity in the model is of prominent importance since the presence of heteroskedasticity in the model means that one of the major assumptions of the standard statistical procedure (assumption of homoscedasticity) is violated, while the presence of structural instability implies that parameter estimates take on

[2] Although they all have similar consequences on statistical analyses (i.e. autocorrelated residuals), note that, strictly speaking, spatial autocorrelation, spatial dependence and spatial association are *not* rigorously identical concepts (Anselin, 1992; Dormann et al., 2007).

different values across distinct geographic areas (also referred to as 'spatial regimes'). Another reason for accounting for the presence of spatial heterogeneity in the model is that such a spatial effect may also be at the root of residual spatial dependence. It is hence likely that spatial autocorrelation diagnosed by Moran's I is produced by an undiagnosed and unmodelled form of spatial heterogeneity (Anselin, 1988; Brunsdon et al., 1999; Le Gallo, 2004)[3]. Throughout the literature, White, Breusch-Pagan and Koenker-Bassett tests are commonly used to test for the presence of heteroskedasticity in the model, while (spatial) Chow tests are used to test for the presence of structural instability in the parameters. Whenever detected, heteroskedasticity can generally be corrected using the 'White correction', which is also called 'Huber-White correction' (Long and Ervin, 2000). Regarding structural instability, the estimation can be carried out either by performing a spatial regime regression (which consists of a regression with varying estimates across '*discrete*' spatial subsets of data), or by applying a Geographically Weighted Regression (GWR) (for which parameter estimates are assumed to vary '*continuously*' across space, i.e. as a function of the latitude and longitude) (see Fotheringham et al. (2000) for further details).

Throughout the literature on *cycle commuting* (and mode choice in general), there is to our knowledge still no or little scientific attention that has been devoted to the presence of spatial autocorrelation and spatial heterogeneity in the data, despite the fact they may lead to statistical biases in the results. It also turns out to be true as regards cycling accidents, although changes are expected to occur over the next few years since more and more attention is devoted to spatial effects in traffic accident research. Examples of applications conducted at the areal level and accounting for spatial autocorrelation and/or spatial heterogeneity indeed appear in ever greater numbers in the literature (see e.g. Miaou et al., 2003; Flahaut et al., 2004; Aguero-Valverde and Jovanis, 2006; Eksler and Lassarre, 2008; Quddus, 2008).

Contrarily to the vast body of literature, this thesis then aims at paying greater attention to the spatial effects by implementing the above mentioned statistical tests and by performing the appropriate statistical models (if necessary). As suggested above (Section 1.2.2) and further in this thesis (Chapters 2), it is *a priori* expected that spatial patterns relative to cycle commuting and accident

[3] If spatial dependence disappears after the spatial heterogeneity has been taken into account, the unmodelled structural instability and/or heteroskedasticity probably caused the observed spatial autocorrelation. Performing a model accounting for spatial autocorrelation is consequently not required any more. However, if spatial autocorrelation persists, the final specification should account for both spatial heterogeneity and spatial autocorrelation. As suggested by Le Gallo (2004) and Anselin (2007), this could be simply achieved by incorporating spatial regimes in e.g. autoregressive models.

risks are strong in Belgium. The importance of cycling policies (including actions, investments, etc.) and the popularity of bicycle use for utilitarian purposes are indeed quite dissimilar between the northern and southern parts of the country. At the network level, exploratory analyses conducted in Chapter 4 also indicate that traffic accidents tend to occur in greater numbers in the vicinity of specific locations (e.g. at intersections), which suggests that spatial clusters of cycling accidents are likely to occur at definite places along the road network. The spatial dimension of the data is then expected to play some role in influencing the results. As a consequence, the potential spatial effects that could result from these data are monitored and, when necessary, are taken into account further in this thesis through innovative spatial approaches (in Chapters 3 and 5). Importantly, these latter in turn allow providing *thorough* results and policy recommendations with regard to pro-cycling strategies and safety measures.

1.5.2.4 Network phenomena and planar assumption

The bulk of empirical studies aiming at exploring and/or explaining the spatial distribution of traffic accidents are based on a planar assumption, i.e. they assume that the real world is represented by a plane. As an illustration, planar point pattern analyses – such as kernel density estimations (KDE) – are widely used in traffic accident research to detect and analyse 'black spots' of accidents over a planar space. Smooth density surfaces are then obtained and mapped over a two-dimensional homogeneous Euclidean space (see e.g. Steenberghen et al., 2004; Pulugurtha et al., 2007; Delmelle and Thill, 2008; Erdogan et al., 2008; Anderson, 2009). However, road accidents are point events constrained to occur on a network space (which is here referred to as a 'one-dimensional space'). In such a case, the assumption of planar space is no longer valid as distances are Euclidean instead of being network-based. Applying planar methods to network-constrained events may indeed lead to biased estimates, which in turn may invalidate the conclusions drawn from point pattern analyses (Yamada and Thill, 2004; Okabe et al., 2006a, 2006b; Xie and Yan, 2008; Okabe et al., 2009). In the light of these issues, Okabe and Yamada (2001) and Okabe et al. (2009) extended several planar methods to a network space, assuming that point events are constrained to a network and that the distance between two of these points is network-based (instead of being Euclidean). In this thesis, Chapter 4 took advantage of such extensions to explore the spatial patterns of cycling accidents along the Brussels' road network. Recent advances implemented in Geographic Information Systems (GIS) by Okabe et al. (2006a; 2006b; 2009) turned out to be helpful in this respect.

1.5.2.5 Estimation of the accident risk at the network level

Strikingly, little is known about the factors that affect the risk of being involved in a traffic accident, although planners and policy makers are generally interested to know which locations and infrastructures are associated with the highest accident risks for all road users. On the one hand, the lack of detailed data on accidents and trip characteristics associated with the different modes of transport (e.g. traffic flows) often precludes researchers to get in-depth knowledge about such risks. As described in Section 1.5.1.2, traffic accident research then either aims at predicting the *accident frequency* or attempts to explain the probability of falling into one level of *injury severity*, as a function of several independent variables/factors (Noland and Quddus, 2004; Lord and Mannering, 2010). This however leads to several well-known methodological issues (e.g. over- or under-dispersion of accident-frequency data) and then requires performing proper statistical approaches in order to avoid invalid inferences (Lord and Mannering, 2010). Besides, the implementation of accident-frequency models within a spatial framework also implies that accident data are commonly aggregated over space and/or time (Liu and Jarrett, 2008), which carries the danger to make wrong inferences about individual-level relationships on the basis of results obtained at an aggregated level of analysis. Such a fallacy – known as the *ecological fallacy* (or ecological bias) – may in turn result in erroneous recommendations if not properly taken into account.

On the other hand, *surveys* aiming at estimating such a risk generally raise some questions about their relevance in providing consistent parameter estimates since they often fail to select valid controls (see e.g. Lusk et al. (2011) and attendant comments). Main issues generally concern the choice and the representativeness of the controls (e.g. how many controls should I choose, and are my controls representative of the actual traffic flows?). To our knowledge, there is no research in the literature that addressed such issues in a rigorous way (from a statistical point of view). For instance, there is generally no justification about the choice of a definite number of controls. Also, the spatial aspects related to the selection of controls (e.g. which sampling method/design should I use to select controls, and is this location really 'safe'?) are totally ignored in the literature, although they are expected to play a key role in the estimation of accident risks. Indeed, nothing or little is said about why some control sites (e.g. reference streets) are selected rather than others, and to what extent such sites are spatially representative of traffic flows (or background exposure).

Within the framework of this thesis, it is aimed at providing a methodological framework other than this provided by studies focussed on surveys and accident models. Chapter 5 then aims at implementing a *spatial* modelling approach at the level of *individuals* (i.e. the cycling accidents) in order to identify which locations and road infrastructures carry the highest risk to cause cycling accidents. Interestingly, such an estimation of the accident risk is based on a case-control strategy and requires the generation of controls, i.e. the creation of data reflecting the exposure of the population under study (i.e. the cyclists) to the outcome of interest (i.e. the accident). In parallel, an exhaustive data set is also created in order to carry out a well-founded sampling of the controls as well as to provide predictions of the accident risk for the whole road network.

1.6 Outline of the thesis

The objective of this thesis is to identify the main factors that influence cycle commuting and the risk of cycling accidents, with the intent to provide scientific-based recommendations that encourage cycling and make it safer. From a methodological point of view, this thesis raises several challenges since our individual research questions differ with regard to many features, notably with regard to the type of data (cycle commuting *vs.* accident data), the level of data aggregation (Belgian municipalities *vs.* individual accidents), the methodologies and attendant research gaps (with respect to e.g. the presence of spatial autocorrelation, the assumption of planar space for a network, underreporting of cycling accidents, etc.), and the study regions (Belgium *vs.* Brussels-Capital Region). To adopt a coherent structure, this thesis is then subdivided into several inter-related parts grouping our research questions on the basis of the above mentioned features (Figure 1.3). Considering that the first part is the introduction, the second one then focuses on the general objective (*i*) of this thesis (see Section 1.3), which aims at examining the spatial factors that influence **cycle commuting** at the scale of the Belgian municipalities. This part hence focuses on Belgium and considers the analysis of **spatially aggregated data**, i.e. data associated with areas (municipalities). Afterwards, the general objective (*ii*) is approached in the third part of this thesis. This latter part thus focuses on one of the main barriers to cycling, i.e. the accident risk for cyclists. More particularly, it aims at examining the spatial factors that influence the risk of being involved in a road accident when cycling. In this latter case, it zooms into the Brussels-Capital Region (Belgium) and focuses on the analysis of **individual/point events** (i.e. the reported and/or unreported **cycling accidents**) along a network space. Finally, the fourth part concludes this thesis and delivers a comprehensive package of recommendations intended for planners and policy makers. Figure 1.3 and the following subsections provide more details about the next chapters of this thesis as well as about the individual research questions.

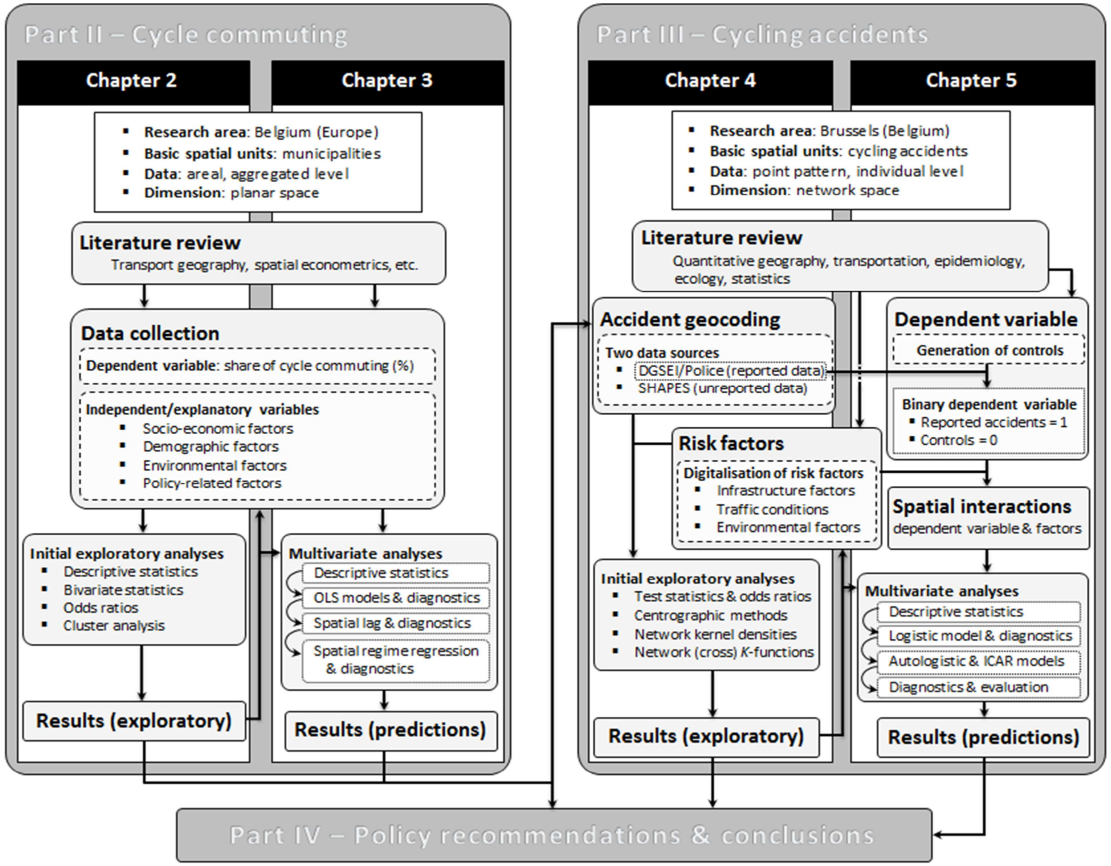


Figure 1.3: General outline of the thesis

1.6.1 Outline of part II – Areal data analyses and cycle commuting (Belgium)

As illustrated in Figure 1.3, this part of the thesis focuses on the spatial factors that influence cycle commuting and use data that are spatially aggregated on Belgian municipalities. It is subdivided into two research questions, or chapters. On the one hand, Chapter 2 mostly makes use of exploratory spatial data analyses and basic statistics to *explore* the spatial factors associated with cycle commuting, as well as the relationship between bicycle use and the risk of accidents for commuters that cycle to work. On the other hand, Chapter 3 adopts a spatial modelling approach to *identify* the significant factors associated with cycle commuting. Here is the detailed outline of these chapters/research questions:

Chapter 2 – Exploratory (spatial) data analysis of cycle commuting and accident risks. This chapter consists of an exploratory step and provides a general overview of cycle commuting and accident risks when cycling in Belgium. It serves as a basis for more robust statistical analyses, complementarily to an exhaustive review of the literature on the factors influencing cycle commuting (carried out in Chapter 3).

The objective of this chapter is twofold. On the one hand, it aims at exploring the relationship between bicycle use and the risk of accidents for commuters that cycle to work in Belgium, while on the other hand it has the objective to examine to what extent urban hierarchy and distance to the workplace influence cycle commuting. This chapter mostly relies on data compiled by the Directorate-General Statistics and Economic Information (2001 census data and accident data for the period 2002-2005). Exploratory analyses of these data are conducted at the scale of the Belgian municipalities and suggest that the variations in cycle commuting are strongly linked to the urban hierarchy, and that high proportions of commuter cyclists are correlated with low risks of becoming seriously injured or killed in a cycling accident. Importantly, our findings exhibit strong spatial differences in cycle commuting and accident risks between the regions, suggesting that spatial factors and/or effects (e.g. spatial autocorrelation, heterogeneity) could play a role in influencing their spatial patterns.

Chapter 3 – Spatial modelling of cycle commuting. This chapter extends the previous exploratory data analyses by explaining the spatial variation of bicycle use for commuting to work at the level of the Belgian municipalities. Within this framework, this chapter aims at identifying which factors significantly influence cycle commuting in order to provide statistically-based recommendations for planners and policy makers. In a first step, we reviewed the factors that influence the use of the bicycle for commuting, before collecting these from a wide range of sources (DGSEI, Federal Police, FPS Mobility and Transports, etc.). In a second step, descriptive statistics and bivariate analyses are computed in order to explore the relationships between each of the factors and the dependent variable (proportion of commuting by bicycle, per municipality). Multivariate analyses are then carried out at the scale of all 589 municipalities in order to confirm some of the results obtained in Chapter 2 and with the aim to provide sounder results. Special attention is paid to spatial effects, since previous findings (Chapter 2) tend to suggest that cycle commuting is strongly affected by spatial autocorrelation and heterogeneity (a clear-cut north-south division of cycle commuting was indeed observed at the scale of the Belgian municipalities). Spatial econometric techniques are then reviewed and used to correct for the presence of spatial autocorrelation, after which a disaggregated modelling strategy is adopted for the northern (Flanders) and southern parts of the country (Wallonia and Brussels) to address the presence of spatial heterogeneity. From a methodological point of view, the modelling techniques applied in this chapter highlight the importance of accounting for spatial dependence and heterogeneity. Indeed, spatial autoregressive models, combined with a strategy disaggregated by region, appear to be very powerful in eliminating such spatial effects in the data.

1.6.2 Outline of part III – Point data analyses along a network space and cycling accidents (Brussels)

The third part of this thesis is devoted to the identification of spatial factors that are associated with one of the main barriers to cycling, i.e. cycling accidents. As indicated in Figure 1.3, empirical analyses are conducted on the Brussels-Capital Region, at a disaggregated level of analysis (individual cycling accidents) and over a one-dimensional space (i.e. the Brussels' road network). This part of the thesis is also subdivided into two chapters. On the one hand, Chapter 4 uses test statistics and point pattern methods extended to networks with the aim to *explore* and compare the spatial patterns of cycling accidents registered by the police with those unregistered (by the police) but collected through an open-based online registration survey (SHAPES survey). Such a comparison between reported and unreported cycling accidents would then have the interest to evaluate whether or not official accident databases neglect important information relative to unreported accidents (e.g. with regard to some specific risk factors). On the other hand, Chapter 5 focuses on a spatial modelling approach conducted within a Bayesian framework and based on a case-control strategy to *identify* the factors that significantly influence the risk of cycling accident. Here is the detailed outline of these two chapters:

Chapter 4 – Exploratory (spatial) data analysis of cycling accidents. This chapter aims at exploring and comparing the spatial patterns of reported and unreported cycling accidents. Furthermore, it aims at analysing whether or not these latter have similar locational tendencies with respect to specific road infrastructures (e.g. intersections, roundabouts, etc.). As mentioned above, empirical analyses are conducted on Brussels. In a first step, the literature on spatial point pattern analyses is reviewed as regards both planar and network spaces. In a second step, accident data are collected and geocoded into a GIS using a semi-automatic process, consisting of manually correcting the results obtained through address matching techniques. After an exhaustive review of the literature in traffic accident research, infrastructure factors that are associated with the presence of cycling accidents are digitised into a GIS from a wide range of sources (e.g. orthophotos, maps, etc.). In a third step, comparative statistics, point pattern exploration techniques and network (cross) K-function methods are carried out and – as far as possible – extended to a network space since cycling accidents are inherently network-constrained events. Such exploratory analyses rightly precede modelling techniques applied in Chapter 5 as they provide important clues on how underreporting could affect the model results. Indeed, significant differences in the spatial patterns and (accident-related) factors would be indicative of the fact that unreported and reported cycling accidents occur at different places along the network and hence that relevant variables (i.e. variables having a significant influence in the occurrence of cycling accidents) are probably neglected when focussing only on cycling accidents that are officially reported by the police.

Chapter 5 – Spatial Bayesian modelling of accident risks for cyclists. This chapter extends the exploratory analyses conducted in Chapter 4 and intends to provide an innovative methodological framework to pinpoint locations at risk for cyclists along a road network. The main objective of this chapter is to identify which are the most significant spatial factors (expected to be) associated with the occurrence of cycling accidents in Brussels, and then which locations carry the highest risk to 'cause' traffic accidents for cyclists. Spatial risk factors and accident data are mostly those employed in Chapter 4, with the exception that we only focus on reported cycling accidents. From a methodological point of view, an extensive review of the literature is first carried out in epidemiology, ecology and statistics in order to get knowledge in case-control studies and Bayesian statistics. In a second step, an innovative and rigorous methodological framework is proposed to construct a binary dependent variable (accident, no accident), which will in turn allow modelling the accident risk for cyclists along the Brussels' network. Such a binary dependent variable is created from coupling the geocoded cycling accidents (cases) with a definite number of control sites, or controls (sites where there is supposedly no cycling accident). Control sites are sampled along the 'bikeable' segments of the road network *and* as a function of an exposure variable representing the bicycle traffic. Gravity-based theory was helpful to construct such a variable and allowed estimating the potential bicycle traffic transiting/stopping in each Brussels' statistical ward. In a third step, descriptive statistics are carried out to explore the relationships between the spatial risk factors and the occurrence of bicycle accidents. Multivariate analyses are then conducted within a Bayesian framework to model the risk to be involved in a road accident when cycling on the Brussels' road network (period 2006-2008). As in Chapter 3, special attention is paid to spatial autocorrelation and spatial heterogeneity by using autologistic and intrinsic conditional autoregressive specifications. Sound and innovative results are then obtained as a result of the application of these spatial modelling approaches. Predictions of the risk of having a cycling accident along the network are also computed along the road network and provide a useful tool for planners and decision makers. Such predictions, once mapped, constitute the main innovation of this thesis and are hoped to gain the upper hand to traditional 'hot spot' methods. Interestingly, they turn out to be powerful in identifying road segments where cycling accidents might have been unreported… or might still occur in a near future. They also help cyclists planning the safest possible routes between specific origins and destinations.

1.6.3 Outline of part IV – Conclusions and policy recommendations

This last part aims at summarizing the main findings, contributions and limitations of this thesis, as well as it delivers a comprehensive package of recommendations intended for planners and policy makers. It finally closes this thesis by proposing some leads for future research.

Part II: Spatial analysis of commuter cycling

Chapter 2

Bicycle use and the accident risk for commuters who cycle to work in Belgium: An exploratory spatial data analysis[1]

Outline

This chapter explores the spatial patterns of bicycle use for commuting and the risk cyclists run being injured in a road accident when commuting to work in Belgium. Exploratory data analyses suggest that the observed differences in the use of the bicycle to get to work are strongly linked to the urban hierarchy: commuters are more inclined to cycle in towns and specifically in regional towns (with 25,000 to 120,000 inhabitants). In large towns (more than 200,000 inhabitants), less commuting by bicycle takes place. The relationship between bicycle use and the risk of being seriously injured or killed in a road accident is also studied. A cluster analysis confirms that high proportions of commuter cyclists are correlated with low risks of becoming seriously injured or killed. It also shows that there are strong spatial differences (regional and between different types of towns) in bicycle use and the risk of an accident, which hence suggests that cycling policies should be spatially differentiated.

[1] This chapter is adapted from the following paper: Vandenbulcke, G., Thomas, I., de Geus, B., Degraeuwe, B., Torfs, R., Meeusen, R., Int Panis, L. (2009). Mapping bicycle use and the risk of accidents for commuters who cycle to work in Belgium. *Transport Policy* 16, 77-87. [http://dx.doi.org/10.1016/j.tranpol.2009.03.004]

2.1 Introduction

As mentioned throughout the previous chapter, the promotion of non-motorised modes of transport is increasingly being recognised as an effective way of addressing environmental, health and mobility externalities generated by the growing use of cars and massive periurbanisation. However, several barriers prevent people from cycling: fear of crime or vandalism, bad weather, hills, danger from traffic, social pressure and long commuting distances are some of the most frequently cited deterrents (see Pucher et al., 1999; Rietveld, 2001; Rietveld and Daniel, 2004; Gatersleben and Appleton, 2007; Parkin et al., 2008). Safety concerns and the lack of an adequate infrastructure are – in particular – major hindrances to bicycle use (Pucher et al., 1999; Parkin et al., 2007). Thus, making bicycle use safer is one of the most essential elements in initiating a substantial shift from car to bicycle. It is hence often recommended in the literature that policy makers and planners take steps such as reducing the amount of motorised traffic in urban centres, developing traffic-calming areas, constructing an infrastructure for cycling, and promoting bikepooling (Pucher et al., 1999; Rietveld, 2001; Pucher and Dijkstra, 2003). Such measures reduce the risk cyclists run of being involved in traffic accidents and improve the individuals' overall perceptions of the dangers of cycling (especially as regards the perceived risk of cycling accident). Consequently, they have great potential to encourage more people to cycle for commuting trips. As mentioned in Chapter 1, this could result in a virtuous circle since greater numbers of cyclists on the road improve the safety of all cyclists. Jacobsen (2003) indeed showed that higher levels of cycling (in terms of distance travelled) are correlated with lower rates of fatalities from cycling. In other words, a 'safety in numbers effect' results from such high levels of cycling.

Surprisingly, little attention is devoted to the investigation of spatial patterns associated with such a 'safety in numbers' effect. Yet, initial examination of spatial data could be of great help in exploring how safe/attractive (or unsafe/unattractive) environments distribute over space, as well as in revealing spatial trends and/or associations with potential explanatory variables. For instance, the identification of spatial clusters/groups of (homogeneous) environments could highlight unexpected relationships with specific variables, such as e.g. the cycling policies (which could in turn emphasize which policies are the most effective in encouraging bicycle use and in making it safer). In continuation with the Jacobsen's findings (2003), it is hence questioned here how environments characterised by low rates of cycling accidents and high percentages of cyclists (and conversely) distribute over space. To our knowledge, there still exists no such exploratory data analysis. This chapter then mainly focuses on the relationship between bicycle use and accident risk at the scale of the 589 Belgian municipalities (the smallest administrative units). Complementarily and as a first step before the main objective, it also explores the variation of bicycle use when commuting as a function of (*i*) the level of urban hierarchy (from the largest towns to rural municipalities), and (*ii*) the distance between the residence and the workplace. This preliminary step allows getting more insight about the factors associated with high (or low) levels of commuter cycling, and – as a corollary – possibly also with low (or high) accident risks for cyclists (since, as above mentioned, 'safety in numbers' effects may result from great numbers of cyclists). Lastly, it is also helpful to identify in an explorative way which (explanatory) variables might be of interest to include into the regression models described in Chapter 3.

The present chapter is structured as follows. After describing the data and the area studied in Section 2.2, we analyse the link between urban hierarchies, distances and bicycle use in Section 2.3, and then propose a clustering of the municipalities according to bicycle practice and accident risk (Section 2.4). We end up with a map that pinpoints the municipalities that combine low (or high) proportions of cyclists with high (or low) risks of accidents. Concluding remarks are finally provided in Section 2.5.

2.2 Data sources and studied area

2.2.1 Studied area

As mentioned in Chapter 1, our analyses are conducted in Belgium. This small and highly urbanised European country covers approximately 30 000 square kilometres and has more than 10 million inhabitants. It is subdivided into three institutional regions: the Brussels-Capital Region, the Flemish (Dutch-speaking) Region and the Walloon (French-speaking) Region (Figure 2.1). Belgium has a tight network of towns, dominated by Brussels (more than 1 million inhabitants); the second largest town is Antwerp, which has approximately 500 000 inhabitants. Towns tend to sprawl into their peripheries. This urban spread favours car use and often leads to more and longer commuting trips, which are not convenient for cycling or walking. However, cycle use is still relatively common in Belgium, compared to other industrialised countries, although the rates are well below those in the Netherlands and Denmark. At the European level (EU 15), Belgium is ranked fourth, with a bicycle share of 2.42% (in traveller-kilometres/person/year), and stands out as one of the countries with the highest share of cyclists (Germany: 2.47%; Denmark: 5.48%; the Netherlands: 6.66%) (EU, 2003; Rietveld and Daniel, 2004).

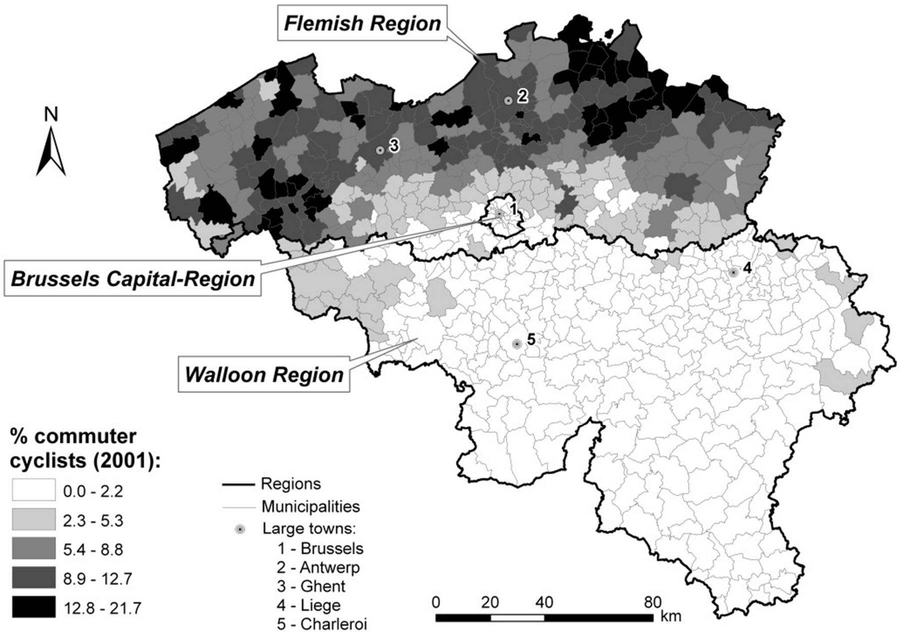


Figure 2.1: Percentage of commuters that use the bicycle as the only mode of transport. Source: DGSEI, 2001

Interestingly, Figure 2.1 shows that strong divergences exist between the northern (Flanders) and the southern part of the country (Wallonia and Brussels). On average, bicycle use for utilitarian purposes is rather common in the north, while it is relegated to a marginal role in the south (mainly recreational activities). As suggested by Chapter 3, such a stark division is explained not only by the culture, but also by a number of political, physical and historical factors (Rietveld and Daniel, 2004; Rodríguez and Joo, 2004). From the 80's, local and regional policies in Flanders played a key role since they early recognised the potential of the bicycle (in terms of sustainability) and paid attention to integrate it in the mobility plans and strategies. Measures favouring cycling – such as the achievement of cycle infrastructures – were hence implemented by the Flemish authorities and contributed to increase (and maintain) bicycle use. Besides this, some physical features also encouraged cycling. Similarly to the Netherlands, Flanders is a flat and highly urbanized region, where most employment is concentrated in town centres. This generates short and, hence, 'bikeable' commuting distances. Also, during the 20's and 30's (and still nowadays), the lack of an extensive public transport system in several Flemish towns[2] probably explained the fact that bicycle use was prefered and historically rooted in the Flemish culture (Albert de la Bruhèze, 1999; Mérenne-Schoumaker et al., 1999; MF, 2002).

Finally, it should also be stated that, in some Flemish municipalities where the university takes up an important place (e.g. Leuven, Gent), the levels of bicycle use in commuting trips are high. This is probably explained by the strong social support generated by the high number of students using the bicycle for their daily journeys (De Bourdeaudhuij et al., 2005; de Geus, 2007).

2.2.2 Data

2.2.2.1 Population census

The 2001 census carried out by the Directorate-General Statistics and Economic Information (DGSEI, 2001) is the most recent database covering the entire population. It not only provides exhaustive information about the demographic, social and professional characteristics of the population, but also gives a large amount of data on mobility (e.g. travel patterns and individuals' mode of transport) and housing characteristics.

The census was used here to compute the **proportion of commuter cyclists**. Interestingly, it reveals that 6.2% of all commuters used the bicycle as their only means of transport between home and workplace, while 68.6% of commuters used a car (Verhetsel et al., 2007). On average, bicycle use was higher in the northern part of the country (Flanders). Indeed, 91% of commuter cyclists live in Flanders (Wallonia: 6.4%; Brussels: 2.6%). The **average total commuting travel time** (return trip) was also taken from the 2001 census and used as a measure of exposure to risk in Equation 2.1 (see Section 2.2.2.2 below). Finally, the **average total commuting distance** (km) was used to analyse the (deterrent) impact of distance on the use of different modes of transport, and more particularly on the use of a bicycle.

[2] Mainly in Western Flanders (e.g. Kortrijk) and in the eastern part of the province of Antwerp (e.g. Turnhout).

2.2.2.2 Road accident statistics

Road accident statistics are compiled annually by the Directorate-General Statistics and Economic Information (DGSEI). They indicate that about 7,200 cyclists were injured or killed in 2002 and almost 8,000 in 2005. However, the number of deaths decreased from 108 in 2002 to 71 in 2005. The data used here are limited to a period of 4 years (2002-2005) and allow the risk of an accident to be computed for each municipality. It is well-known that these statistics strongly underestimate the total number of cycling accidents, particularly when the cyclist is the only person involved and/or when no hospitalisation is involved. In Belgium, several authors have estimated that about 15% of cycling accidents are officially reported (see Doom and Derweduwen, 2005; De Mol and Lammar, 2006; BRSI, 2006). As no correction exists for the entire country, only accidents involving serious casualties (i.e. requiring more than 24h hospital treatments) and fatal accidents were included since these are systematically registered. An index of risk (R_i) was computed and used as a proxy for cyclists exposure to severe/fatal accidents:

$$R_i = N_i / T_i \tag{2.1}$$

where N_i is the average annual number of injuries to cyclists aged between 18 and 65 years, between 2002 and 2005 and occurring on weekdays in municipality i. T_i is the total time (return trip) spent travelling by commuter cyclists living in municipality i per year (assuming 232 working days). It is considered as the exposure time to potential injury from commuter cycling. Note that in municipalities with less than 10 regular cycle commuters, the total commuting time T_i was interpolated from the average times in neighbouring municipalities. More importantly, great care should be paid when analysing R_i as accident data (and thus N_i) do not allow us gaining information about the trip purpose of the cyclist at the time of the accident. In other words, accident data sets do not allow distinguishing between commuting trips and other trip purposes. Although it was here attempted to select accidents concerning commuter cyclists only (through the selection of weekday-related accidents and cyclists aged between 18-65), R_i is likely to be over-estimated. This is even truer in urban areas where the diversity of trip purposes is higher (due to the proximity to several types of facilities, e.g. food shops, leisure areas, etc.), as well as in municipalities where school-related and recreational trips are common during the whole week.

Figure 2.2 indicates that in Flanders, the risk of a cyclist being seriously injured or killed in an accident was spatially homogeneous and lower than the average for the whole of Belgium ($\overline{R}_i = 0.069$, i.e. nearly 7 severe/fatal accidents occur when 10 000 000 bicycle-minutes are achieved). Only a few Flemish municipalities on the coast, near the linguistic border, in Limburg (Flemish provincy, in the north-east) or in the periphery of Brussels had risks of severe/fatal accidents higher than the mean. In Wallonia, the risk of a cyclist being seriously injured or killed in an accident were much more varied: there was a very low risk (equal or close to zero) in the majority of municipalities (due to the fact that very few if any cyclists were seriously injured or killed). On the other hand, nearly 38% of municipalities had quite a high risk of severe/fatal accident.

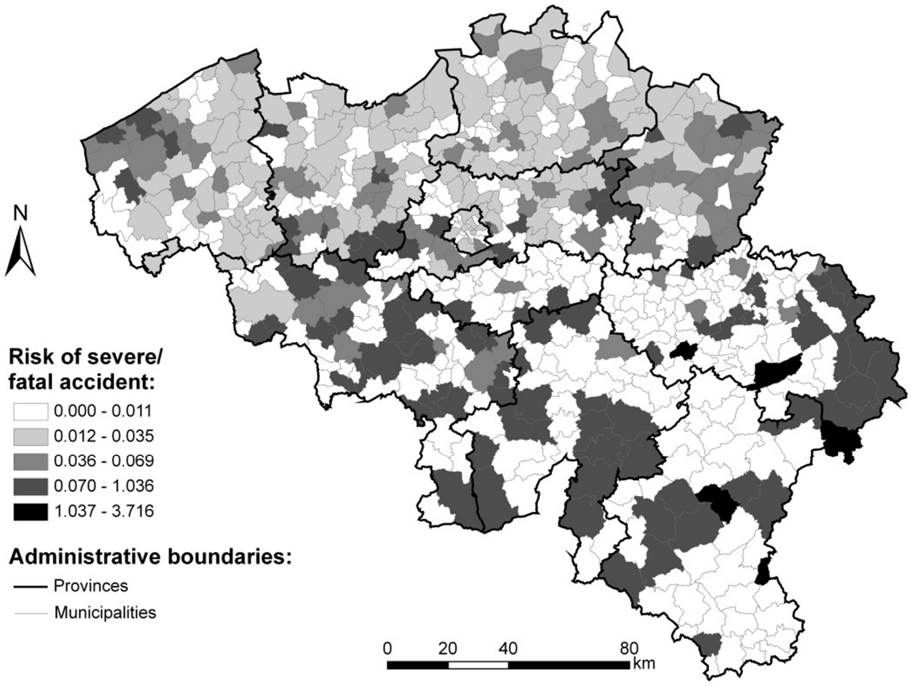


Figure 2.2: Risk of severe/fatal accident, defined as the average number of severe and fatal accidents (for commuter cyclists) per 100,000 bicycle-minutes, by municipality

Interestingly, a low risk of severe/fatal accident was observed in most of the large towns, which seems to suggest that an urban environment is safer than a rural one for commuter cyclists. This may be partly explained by the large number of hurdles (e.g. traffic lights, pedestrian crossings, congestion) that reduce the speed of traffic in towns. However this is not true for all towns: moderate or high risks are observed in some regional towns (25 000 to 120 000 inhabitants).

Table 2.1: The means of variables in municipalities with different ranks in the urban hierarchy (H_j)

Description	Source	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8
% of commuter who cycle	2001 Census	4.65	8.89	7.11	5.22	5.59	4.83	4.73	2.16
Median income (in euro)	DGSEI (2001)	17010	18733	19135	19247	18855	19282	19789	19287
Population density (inhabitants/km²)	DGSEI (2001)	2460	912	945	399	556	1545	342	160
Jobs density (jobs/km²)	DGSEI (2001)	1877.25	374.16	367.58	115.43	146.29	484.53	62.38	31.86
% of economically active people below 25 years of age	2001 Census	10.57	10.90	10.54	10.74	10.22	10.10	9.89	9.39
% of economically active people above 54 years of age	2001 Census	7.71	6.96	7.34	6.91	7.08	7.13	6.84	6.73
% of economically active people having only primary schooling	2001 Census	7.38	6.06	6.12	5.95	6.04	6.17	5.92	5.63
% of economically active people having a school leaving certificate as their highest qualification	2001 Census	52.51	54.22	56.01	58.64	58.56	56.83	57.86	58.11
% of economically active people having a university degree	2001 Census	40.11	39.72	37.87	35.41	35.40	37.00	36.22	36.26
% of households without children	2001 Census	77.64	73.94	73.08	70.69	70.27	70.75	68.43	67.28
% of households that do not own any bicycles	DGSEI (2001)	57.65	35.82	32.95	35.00	33.93	35.60	27.76	33.48
% of households that do not own any cars	DGSEI (2001)	37.78	25.99	22.34	21.09	20.26	21.06	15.56	15.14
% of households estimating they have low-quality cycle facilities in their neighbourhood	2001 Census	68.89	59.59	59.46	66.87	63.68	63.32	63.73	73.82
Average daily commuting distance (kilometres)	2001 Census	17.31	19.40	19.26	22.86	22.05	20.56	22.86	27.02
Annual number of bicycle thefts per 100 cyclists	Federal Police (2000-2002)	15.82	13.78	13.91	13.64	12.16	11.08	6.89	5.31
Average number of severe/fatal accidents (cyclists) per 100,000 bicycle minutes (i.e. total minutes spent commuting by bicycle)	DGSEI (2002-2005) and 2001 Census	0.02	0.03	0.04	0.07	0.04	0.05	0.06	0.13
% of surface area dedicated to public services (e.g. council offices, schools)	DGSEI (2004)	4.52	2.09	1.77	0.87	1.12	1.78	0.53	0.21

continued on next page

continued

Description	Source	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8
% of surface area which is built up	DGSEI (2004)	78.00	45.45	36.67	26.45	30.66	39.95	24.04	14.38
Number of vehicles (million) by kilometre of regional road	FPS Mobility and Transports (DGSEI, 2000)	5.69	4.12	3.87	2.56	3.26	3.79	2.94	1.99
Number of vehicles (million) by kilometre of municipal road	FPS Mobility and Transports (DGSEI, 2000)	0.90	0.46	0.30	0.20	0.23	0.27	0.13	0.08
% of inhabitants declaring they are in a bad state of health	2001 Census	29.44	25.01	24.31	23.60	25.22	24.77	23.29	24.74

2.2.2.3 Urban hierarchy

Ranks are attributed to the municipalities on the basis of an index calculated by Van Hecke (1998) and based on the degree of equipment of the municipality as well as on its attractiveness. The degree of equipment was calculated using both the quantitative (e.g. number of hospitals) and qualitative importance of the facilities (e.g. presence of universities), while the attractiveness was estimated on the basis of the visitor flows attracted by these facilities. They are denoted by H_j (j = 1,..., 8; see Appendix A.1) and range from H_1 for the largest towns (more than 200 000 inhabitants; e.g. Brussels or Antwerp) to the smallest and least-populated municipalities H_8 (rural municipalities).

Table 2.1 lists some of the socio-economic and environmental features of each rank. In particular, it indicates that population and job densities as well as urban land use are high in municipalities in the first three ranks of the hierarchy (H_1 to H_3). The opposite situation is true for rural municipalities (H_8). This to a large extent explains the differences in the commuting distances between towns (where the proximity of different activities is high) and rural areas: the shortest average commuting distances are found in the largest towns. Finally, high traffic volumes are observed along the municipal and regional road networks in urban municipalities. The large number of activities (e.g. jobs, leisure, public services) and inhabitants make such municipalities highly attractive, leading to high traffic densities.

2.3 Bicycle use *versus* urban hierarchy

2.3.1 Background

Commuting distance is often considered to be one of the main deterrents to bicycle use and it is closely related to the level of urbanisation. Only people living close to their workplace (less than 10 km) even consider cycling to work (Kingham et al., 2001; Bergström and Magnusson, 2003; Dickinson et al., 2003; Saelens et al., 2003; Pucher and Buehler, 2006; Parkin et al., 2008; Verhetsel and Vanelslander, 2010). The distances commuters (would be likely to) cycle depend on land-use and transportation features but also on a large range of socio-economic and demographic factors, and the physical and weather conditions. For instance, commuters faced with steep slopes and/or strong wind speeds will only cycle short distances or will simply not consider cycling (Rodriguez and Joo, 2004; Rietveld, 2001; Rietveld and Daniel, 2004). Many research papers also show that age and gender are determinants of trip distances and bicycle use. For example, Dickinson et al. (2003) showed that women in the UK cycle less, and make shorter commuting trips than men (due to factors such as personal security reasons and family commitments). As shown in Figure 2.3 (as well as in Chapter 3), this is also true for Belgium but only up to a certain age. Women over 60 cycle longer commuting distances than men of the same age. This could be because they are fitter (Deboosere et al., 2006), or it could be because of the relatively low proportion (50-70%) of women in this age group who have a driving license (Hubert and Toint, 2002). As well as age and gender, other factors affecting the (un)willingness of employees to cycle long distances to work are the provision of appropriate facilities at the workplace (e.g. cycle lockers, changing facilities) and the dress code imposed by their company (Dickinson et al., 2003).

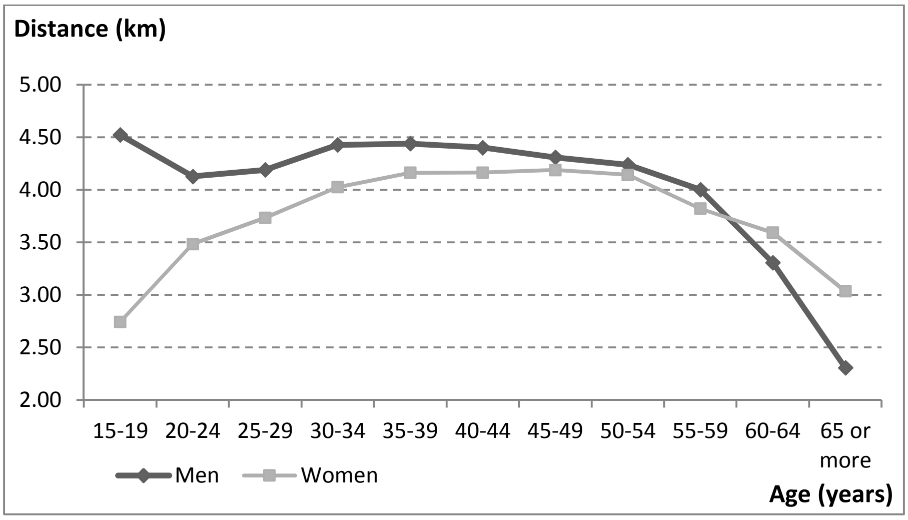


Figure 2.3: Age and gender of cyclists versus commuting distance (one-way)

From a planning point of view, compact and mixed-use environments make it easier to undertake specific activity schedules (e.g. work, recreational activities) by bicycle. Trip distances tend to be shorter and more bikeable in these areas, owing to the close proximity of different places and facilities (Cervero and Kockelman, 1997; Kitamura et al., 1997; Meurs and Haaijer, 2001; Saelens et al., 2003; Pucher and Buehler, 2006). The proximity of public transport interchanges in urban areas (e.g. railway or metro stations) also makes it possible to use a bicycle as a complementary mode of transport, and to combine it with public transport (Martens, 2004; Martens, 2007). Nevertheless, the extensive provision of public transport in urban areas also makes that mechanism highly attractive and competitive, even for distances which are considered feasible for cycling. This probably explains the fact that, in such areas, public transport is used more intensively than bicycles (Ortúzar et al., 2000).

2.3.2 Exploratory data analysis

2.3.2.1 Cycling, urban hierarchy and distances

The previous section raises the following questions: do the largest Belgian towns favour bicycle use? Are they the most favourable environments for cycling? Also, do the increasing commuting distances have a deterrent impact on bicycle use? In order to get right to the bottom of such questions, several exploratory data analyses were performed on the census data and aimed at comparing the bicycle share with urban hierarchy and commuting distances in the same figures.

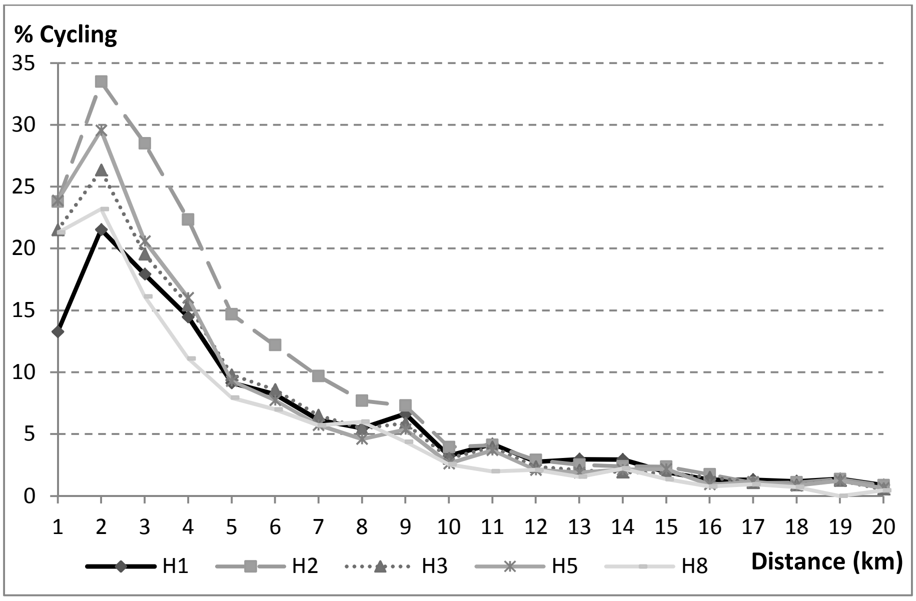


Figure 2.4: Proportion of commuters that cycle in function of the urban hierarchy (H_j) of the workplace and commuting distances (2001)

Figure 2.4 shows the proportion of commuters who cycled (Y-axis) as a function of the distance they travelled to work (X-axis) and the type of municipality in which their workplace was situated. It confirms that, for most people, 10 kilometres is the limit for cycling to work, whatever the environment of the workplace. Below that limit, bicycle use is more frequent in urban environments.

However, the rank of the town also plays an important role: for distances below 5 km, cycling appears to be most popular in municipalities of rank H_2 (regional towns), while large towns (H_1) are characterised by the lowest proportion of cycling commuter (only approximately half the rate in H_2). This can be explained by the fact that in large towns (H_1), walking is frequent due to the close proximity of different places/activities. Public transport is also well-developed (e.g. dense network, high frequency, comfort) and hence is highly competitive to cycling (Figure 2.5). It encourages intermodal journeys including walking as feeder mode (for both access and egress trips). In large towns such as Brussels, the distance between the place of residence (or work) and the closest public transport stop/station is generally short: approximately 96% of the inhabitants (and jobs) are located less than 500m from the closest public transport stop (Vandenbulcke et al., 2007). In H_1 towns, traffic is also much denser than elsewhere, and an adequate cycle infrastructure is often lacking. The high population densities observed in large towns (Table 2.1) may also play a role in the sense that households living in large towns generally do not have any garage (especially in the densest areas) and have probably little room in the flat to store their own bicycles. All these reasons may dissuade people from cycling to work in large towns.

In smaller towns (H_2), road traffic is less dense and public transport is often limited to buses (no tram or metro). This may explain the popularity of the bicycle for commuting. Moreover, many of these regional towns (e.g. Bruges, Leuven) are located in Flanders where cycling is traditionally more common and where the town networks are tighter. In Wallonia, the town networks are looser, leading to longer (and hence unbikeable) commuting distances (see Verhetsel et al., 2007).

Figures 2.5 and 2.6 confirm that the proportion of commuters who walk to work is very high (60%) for trips of less than 1 kilometre, but decreases sharply with increasing distances. By contrast, the proportion travelling by car increases steadily with distance: it is more than 40% for trips of 2 kilometres and rises to over 60% for distances of 5 kilometres or more. Comparing Figures 2.5 (H_1) and 2.6 (H_2) confirms the importance of the size/rank of the destination town on mode of transport: bicycle use is greater in H_2 than in H_1, after which it slightly decreases for smaller towns[1]. Fewer commuters use public transport to reach low-ranked municipalities (e.g. H_8), probably because of the poorer quality of public transport (which also explains the high figures for car use in such areas).

[1] The figures for smaller towns and rural areas are not shown here.

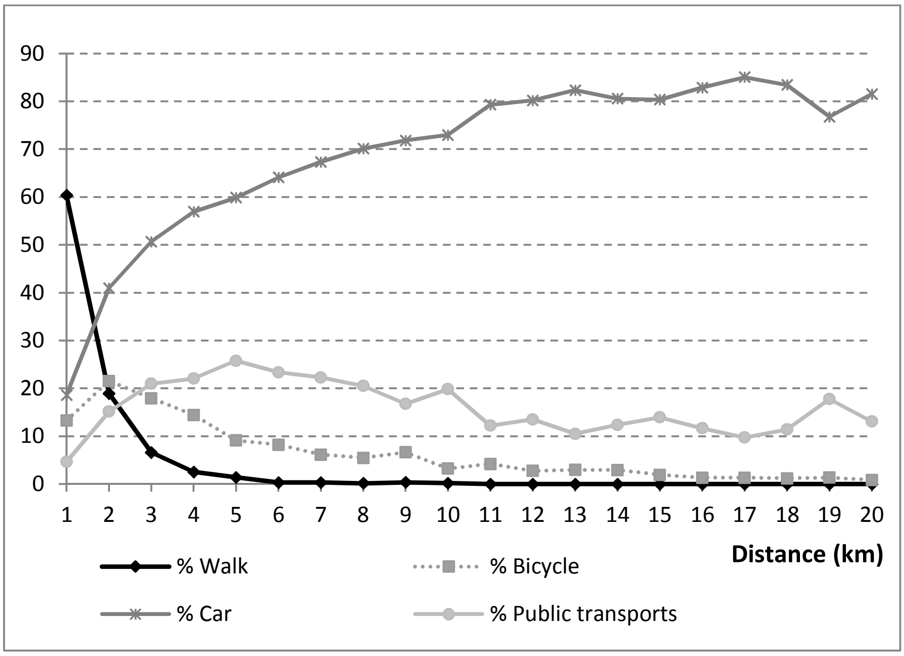


Figure 2.5: Modal share for towns H_1 as destination – commuting trips (2001)

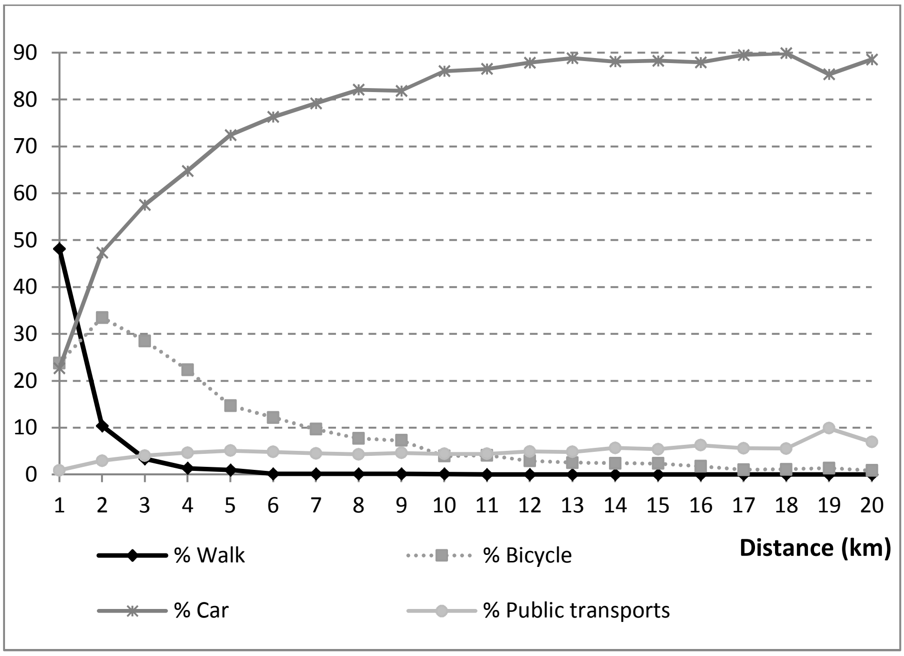


Figure 2.6: Modal share for towns H_2 as destination – commuting trips (2001)

2.3.2.2 Odds ratios

Odds ratios (*OR*s) were computed to compare the probability of commuting by bicycle to one type of destination to the probability of commuting by bicycle to another. In other words, we compared the likelihood of commuting by bicycle to different hierarchical ranks of urbanisation (used as destinations). The *OR* method has several advantages (symmetrical, convenient mathematical properties, easy to interpret) that makes it a good measure for comparing the relative likelihood of an event in two groups (for further information, see: Daya, 2000; Simon, 2001; Prasad, 2007).

Let us compare H_2 municipalities with the other ranks of urbanisation (Figure 2.7). For a pair of ranks such as H_2 and H_1, we compute the *OR* as:

$$OR = \frac{\textit{odds} \text{ of cycling towards } \mathrm{H}_2 \text{ municipalities}}{\textit{odds} \text{ of cycling towards } \mathrm{H}_1 \text{ municipalities}} = \frac{a/c}{b/d} \tag{2.2}$$

where a is the number of cyclists commuting to H_2, b is the number of cyclists commuting to H_1, c is the number of 'non-cyclists' commuting to H_2, and d is the number of 'non-cyclists' commuting to H_1. *OR*s greater than 1.0 indicate that the event (cycling) is more likely in the first group (or rank). Conversely *OR*s below 1.0 indicate that the event is more likely in the second group.

From Figure 2.7, we see that the odds of commuting by bicycle to regional towns (H_2) compared to other types of destination increase for journeys of up to 3 or 4 km. Above 3 or 4 km, there is a decrease but the odds are still greater than 1.0. This suggests that large distances progressively offset the features that make H_2 municipalities more 'bikeable' than other types of municipalities. This confirms the results shown in Figures 2.5 and 2.6. The only exception to the rule is when the regional towns are compared to H_1 (large towns): the odds of commuting by bicycle are higher and constantly decrease up to 14 km (instead of increasing up to 3-4 km). This could be explained by the high proportion of commuters who walk and/or take the public transport in the large towns (which is due to the high proximity and the presence of an extensive public transport system, here). Also, the observed decrease of the odds for journeys of up to 3 or 4 km is probably explained by the slighter drop of bicycle use in H_1 towns.

Finally, the *OR*s are generally higher than 1.0 for H_2 destinations, whatever the distances and the rank of the municipality with which H_2 is compared. Commuters are more likely to travel by bike to H_2 municipalities than to other types of municipalities. As an illustration, let us consider a commuting distance of 5 km. In this case, there is a 1.71-fold (95% confidence intervals: 1.63 to 1.79) greater chance of commuters cycling to work in a regional town (H_2) than in a large town (rank H_1).

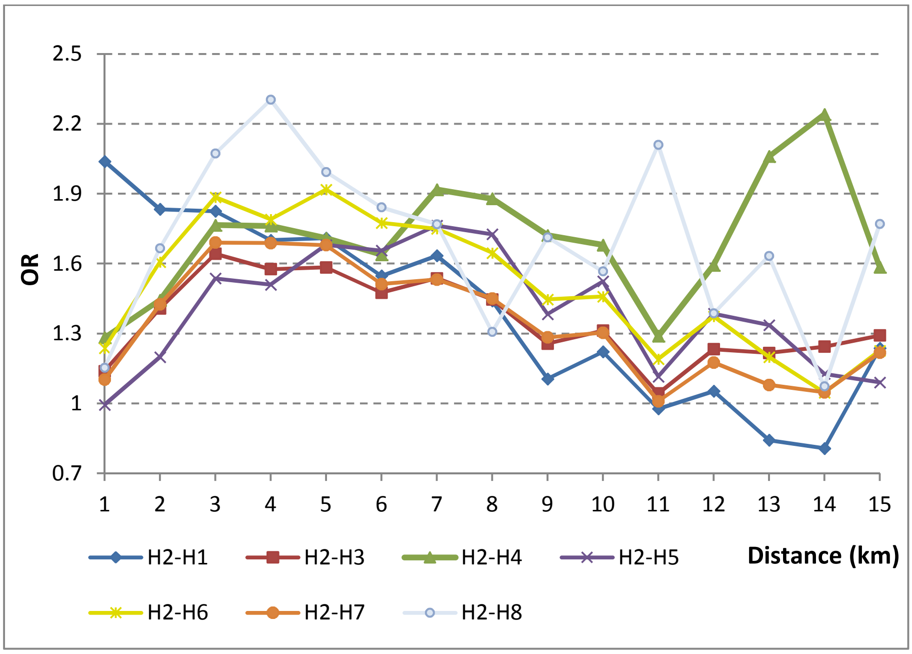


Figure 2.7: Odds ratios for cycling (H_2 vs H_j) as a function of the travelled commuting distance. For legibility reasons, confidence intervals are not reported here.

2.3.2.3 Exploring inter-municipality differences

In this section, we aim to identify some of the factors that make some urban ranks more bikeable than others. In particular, we focus on the differences between H_2 (more bikeable) and H_1 (less bikeable). The descriptive statistics and correlation coefficients shown in Tables 2.1 and 2.2 suggest that the combination of several features (high job and population densities, low car availability, small commuting distances) explains the fact that H_2 municipalities are more bikeable. In particular, the greater provision of public services (such as schools and hospitals) probably generates a lot of cycling trips since a relatively high proportion of commuter cyclists (38%) work in the public sector (DGSEI, 2001). From a safety-related point of view, H_2 areas are also characterised by low risks of being seriously injured or killed and better cycle infrastructures than in rural areas.

Table 2.2: Spearman and Pearson correlation coefficients between some selected variables (expected to be explanatories) and bicycle use as well as urban hierarchy (n = 589)

Description of variables	Source	Correlation with urban hierarchy‡	Correlation with bicycle use‡
Urban hierarchy of municipalities (largest towns = 1; smallest villages = 8)†	KUL (1998)	–	**–0.23**
% of commuter cyclists†	2001 Census	**–0.23**	–
% of active people less than 25 years of age	2001 Census	**–0.19**	0.54
% of households without children	2001 Census	**–0.48**	0.23
% of households that do not own a car†	DGSEI (2001)	**–0.45**	–0.25
% of households that do not own a bicycle†	DGSEI (2001)	**–0.11**	–0.85
Average commuting distance	2001 Census	**0.34**	–0.54
Population density (inhabitants/km²)	DGSEI (2001)	**–0.50**	0.28
Job density (jobs/km²)	DGSEI (2001)	**–0.62**	0.38
% of area used for public facilities (e.g. council offices, schools)†	DGSEI (2004)	**–0.62**	0.17
% of households estimating they have low-quality cycle facilities in their neighbourhood	2001 Census	**0.16**	–0.82
Annual number of bicycle thefts for 100 cyclists†	Federal Police (2000-2002)	**–0.47**	n.s.
Annual number of severe/fatal accidents (cyclists) per 100,000 bicycle minutes†	DGSEI (2002-2005) and 2001 Census	**0.25**	–0.20

† Logarithmically transformed variables (*ln(x+1)*)
‡All correlation coefficients significant at the level 99.9%
n.s.: not significant at the 90% level
Normal font: Pearson product moment correlations; **Bold font**: Spearman correlations

Similar features occur in the largest towns (H_1) but the greater volume of motorised traffic discourages cycling and makes H_1 destinations less attractive to cyclists than H_2 (which will be confirmed in Chapter 3). Moreover, large towns (H_1) are characterised by high proportions of households that do not own a bicycle: more than 54% of households in Brussels and in the Walloon towns (e.g. Charleroi, Liège) do not possess a bicycle (DGSEI, 2001). The risk of bicycle thefts is also high in these areas (Banister and Gallent, 1998; Rietveld, 2001; Rietveld and Daniel, 2004). Tables 2.1 and 2.2 show that there is a negative correlation (–0.47) between the annual number of bicycle thefts per 100 cyclists and the rank in the urban hierarchy, suggesting that the risk of bicycle theft is greater in urban areas. Since bicycle use is lower in the largest Belgian towns, we suspect that bicycle thefts may have a deterrent impact on cycling in such areas. However, this cannot be confirmed because the relationship between the risk of bicycle theft and the likelihood of cycling to work is not significant (Table 2.2).

2.4 Bicycle use and risk

After having identified the places where cycling is the most frequent for commuting trips (Section 2.3), we now analyse the relationship between bicycle use and risk of severe/fatal accidents, and then explore the spatial variation of this relationship. Do the municipalities with high (low) bicycle use and low (high) risks of severe/fatal accident concentrate in space, and why? Are there regional differences in terms of cycling policies or driving behaviour? Road accident statistics (DGSEI) and the analyses performed in Section 2.3 suggest some likely causal factors leading to the observed spatial patterns (clusters).

2.4.1 Clustering municipalities

Table 2.2 shows that – as expected – the risk of cyclists becoming seriously injured or killed in a road accident decreases as the proportion of cyclists increases ($r = -0.20$ with a log transformation of both variables; r is significant at the 99.9% level). This confirms previous analyses (e.g. Wardlaw, 2000; Jacobsen, 2003; Pucher and Buehler, 2006). To explore this topic further, we clustered the 589 Belgian municipalities according to the risk of bicycle accidents (R_i) and the proportion of cyclists among commuters, using Ward's ascending hierarchical method (Ward, 1963). At each step, this method minimises the sum of squares of any pair of clusters to be merged, so that the two closest clusters are joined to form a new cluster. In order to determine the optimum number of clusters, the *CCC* (cubic clustering criterion), the pseudo-F statistic (*PSF*), the pseudo-t^2 (*PST2*), the semi-partial R-squared (*SPRSQ*) and the R-squared (*RSQ*) were helpful (see Fernandez, 2002; Tufféry, 2005 for further information). These statistics suggest the use of eight clusters for the classification. The results help us to understand the geography of road accidents for cyclists, and suggest clues for local policy.

Figure 2.8 shows interesting spatial patterns, and emphasises the regional differences. Municipalities in clusters A, B and C provide the most 'bikeable' environments (i.e. high and safe bicycle use) while those in clusters F, G and H are regarded as the least bikeable (i.e. low and unsafe bicycle use). The map also indicates that the most and least bikeable environments spatially cluster, so leading to a clear-cut north-south division (positive spatial autocorrelation[2]). Such a division could be indicative of the fact that different (regional) policies are implemented in terms of bicycle promotion and safety.

2.4.2 Analysis of the results

Clusters A, **B** and **C** all have a low or moderate risk of severe/fatal accident for cyclists, combined with moderate or high proportions of commuter cyclists (a few municipalities located on the coast or in Limburg have moderate accident risks, but the high use of bicycles offsets these risks). Such municipalities – mainly located in Flanders – are characterised by a safe and attractive environment for cyclists, encouraging cycling and leading to a virtuous circle (since more cyclists on roads may reduce the risk of cyclists having accidents). In such municipalities, the availability of an adequate cycle infrastructure (e.g. cycle lanes, traffic lights for cyclists at junctions), the flat terrain, the lifestyle, and the presence of pro-cycling policies are some of the factors that stimulate cycling (Rietveld, 2001; Rietveld and Daniel, 2004; Witlox and Tindemans, 2004) and – as a consequence – make it safer through the presence of a 'safety in numbers' effect. Also, many of the car drivers living in the Flemish Region are themselves cyclists (when commuting trips or for other purposes) and are perhaps more respectful towards commuter cyclists than drivers elsewhere.

Cluster D covers municipalities that have both a small proportion of cycling commuters and a very low risk of cyclists becoming seriously injured or killed (equal or close to zero). They are mainly located in Wallonia and consist of urbanised and rural municipalities. In most of these municipalities, there were no severe/fatal cycling accidents during the period studied (2002-2005) (in other words, the risk of severe/fatal accident was zero). However, in three municipalities (Uccle, Namur and Liège), the risk of being seriously injured or killed was not zero, although it was still low. These municipalities are all highly urbanised, suggesting that the numerous impediments, such as crossroads and pedestrian crossings, slow down the faster traffic and so decrease the danger for cyclists.

[2] This result could be related to the distribution of the two clustered variables (bicycle use and casualty risk), which is far from being random. Moran's I statistics are 0.90 and 0.13 ($p <$ 0.0001) for 'bicycle use' and 'casualty risk' variables, respectively.

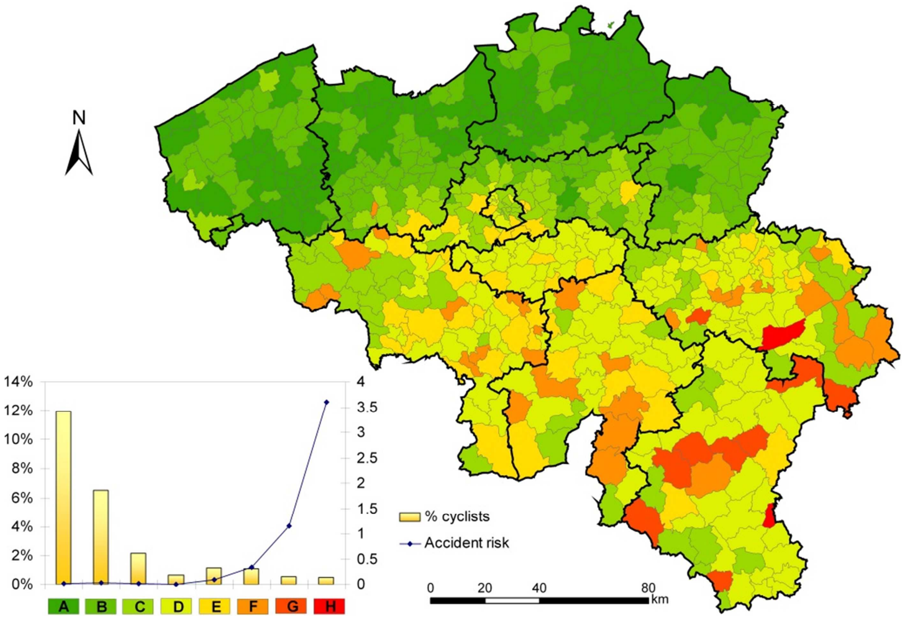


Figure 2.8: Classification of municipalities based on the two variables: bicycle use and risk of severe/fatal accident

Moderate bicycle use together with moderate or high risks of severe/fatal accidents is found in **clusters E** and **F**. Most of the municipalities included in these clusters are in Wallonia, although a few are in areas of Flanders close to Brussels. Every day, a large amount of traffic converges on Brussels, having passed through neighbouring municipalities, which may well increase the risk to cyclists in these latter. Road accident statistics (DGSEI) seem to confirm this assumption, in that they show that the proportion of accidents involving motorised vehicles is high in these municipalities. One of the main factors triggering accidents is the driving behaviour of motorists: car drivers often make bad manoeuvres or lose control of their vehicle (*ibid.*). Moreover, they frequently do not respect the right-of-way (*ibid.*), which illustrates both the fact that cyclists are not an integral part of the 'street scene' and that motorists do not always respect cyclists (especially in Wallonia). Cyclists constitute only a low proportion of the road traffic (i.e. they have low visibility) and most road users have never themselves experienced cycling as a way of commuting, which suggests that they cannot really put themselves in the cyclist's place. Of course, some accidents are caused by the cyclists themselves, when they do not follow the traffic signals (right-of-way) or are not in the correct place. Many accidents also happen when cyclists lose control of their bike or simply fall. Surprisingly, few of these accidents are caused by bad weather (e.g. rain or snow) and/or bad road conditions (e.g. wet or dirty roads), which suggests that other reasons are at the root of the accident: e.g. driving a poorly maintained bicycle or performing a wrong manoeuvre. The prevalence of such accidents suggests that improving cycle infrastructure and traffic education in some parts of the country (especially in Wallonia and Brussels) might help to reduce the risk of severe/fatal accident among cyclists, as well as the overall perception of this risk.

Finally, **clusters G** and **H** consist of hilly municipalities, characterised by high rates of severe/fatal accidents and low proportions of cycling commuters. The commuting distances in these municipalities are generally large and the road network is often winding and sometimes steep. They constitute unsafe and, consequently, unattractive environments for (potential) cycling commuters. All the municipalities in these clusters are located in Wallonia. According to the road accident statistics (*ibid.*), most of the serious injuries and fatalities there are due to the fact that riders fall off their bike, which may suggest that the lower visibility on the roads (due to winding roads) is one of the factors that play a role in the accident occurrence. Some accidents also happen when the motorised vehicles overtake the cyclist or do not respect his or her right-of-way (*ibid.*). The driving of motorists and the fact that cyclists are unusual on Walloon roads probably explain such occurrences. The lack of a high-quality infrastructure may also play a role: the 2001 census indicates that more than 80% of households in these municipalities are not satisfied with the state of the cycle paths there (DGSEI, 2001; Verhetsel et al., 2007), which suggests that the accidents are not inevitable.

2.4.3 Regional and inter-municipality differences

Strong regional differences then exist between Flanders and Wallonia in terms of bicycle use and risks of severe/fatal accident. As it is the case in countries such as Denmark or the Netherlands (Rietveld and Daniel, 2004), cycling in Flanders is part of the lifestyle and benefits from a cultural tradition (Toint et al., 2001). Cyclists are here perceived as legitimate road users and motorists are generally mindful and respectful towards them, especially because many cycle themselves (Pucher et al., 1999). Such an environment then results in a better road safety and encourages cycling. In contrast, in Wallonia, bicycle use in commuting trips is rather marginal. Commuters living here face barriers such as hilly terrains, lack of adequate cycle facilities and large commuting distances. Principal Component Analyses (not illustrated here) and results in Chapter 3 confirm these statements and suggest that such barriers discourage cycling, whereas the presence of high-quality cycle facilities, flat terrain and short/moderate commuting distances encourage it in Flanders.

Differences between the Regions are confirmed in Figure 2.9. It shows that, whatever the rank in the urban hierarchy, bicycle share in Flanders is substantially higher than in Wallonia. Furthermore, the risk of severe/fatal accident is low and differs very little from rank to rank in Flanders, while it varies a lot in Wallonia. Strong differences are also observed between the different ranks and confirm some previous findings (Section 2.2.2): the highest bicycle shares are observed in regional towns (H_2) whereas the lowest ones are found in rural municipalities (H_6-H_8). As previously mentioned, factors such as moderate or high population and job densities as well as the mixity of land-use explain such bicycle shares in regional towns. Note however that the distribution of bicycle use is somewhat different when the analysis is performed at a regional level. Indeed, the Flemish Region reveals a first peak of bicycle share for H_2 and a second for H_5, while the Walloon Region shows only one peak for H_3.

Another interesting result is the fact that high bicycle shares are correlated with low rates of severe/fatal cycling accidents. The results indeed show that Flemish municipalities are characterized by high bicycle shares and low risks of being seriously injured or killed for a cyclist in a road accident, while the Walloon municipalities generally show the opposite trend (low bicycle use and moderate-high risks of severe/fatal accident). It then suggests that risks of severe/fatal accident for cyclists are closely related to the level of bicycle use. According to Jacobsen (2003), the improved safety of cyclists is mainly explained by a behaviour modification of motorists, caused by the great numbers of cyclists in traffic: motorists tend to adapt their driving behaviour when they expect cyclists or experience cycling themselves.

Finally, Figure 2.9 also shows that the risks for cyclists of being seriously injured or killed in an accident increase when the environment is less-urbanized, suggesting that the risk of being seriously injured or killed is higher in rural municipalities. Whatever the region, the greater number of 'physical barriers' (e.g. traffic lights, road humps) as well as the congestion, the lower speed limits and the higher shares of walking trips in urban areas force motorists to slow down and adopt a more careful driving behaviour. Such factors then reduce the differential speed between fast and slow modes and, consequently, decrease the number of severe/fatal accidents. This last assumption is confirmed in a study conducted by Daniels and Geurts (2004) on traffic unsafety in the Flemish Region. They noted that less-urbanized environments are characterized by high speed limits (and effective driving speeds) and then have a higher number of serious injuries and fatalities in traffic accidents. Interestingly, Verhetsel et al. (2007) also observed high driving speeds in such less-urbanized areas (especially in Wallonia) when they analysed the 2001 census data.

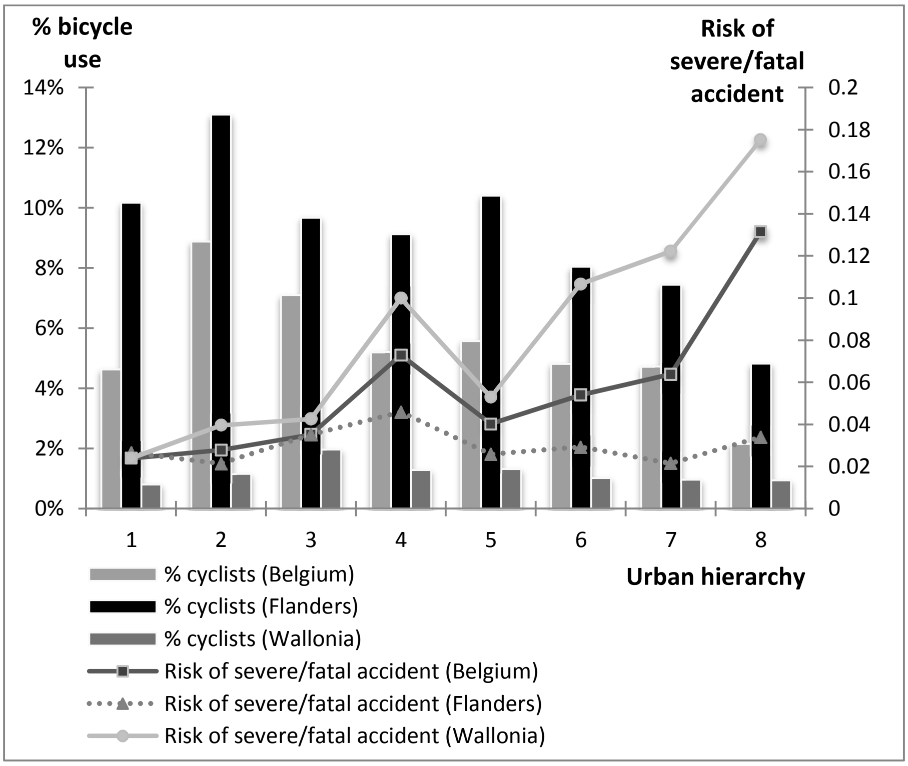


Figure 2.9: Modal share of cyclists *versus* risk of severe/fatal accident: regional differences. Brussels is not illustrated here because of the small number of municipalities (19)

2.5 Conclusion

This chapter has explored the relationship between commuting by bicycle and accidents to cyclists, as well as the extent to which urban hierarchy and distance to the workplace influence bicycle use. We have shown that, in Belgium, urban environments encourage the use of non-motorised modes of transport and more particularly cycling. The presence of a densely built-up environment generates short commuting distances and hence encourages cycling. At the opposite extreme, commuters who live in low-density areas usually have to cover longer distances to work, and consequently depend more on motorised transport (especially private cars) since public transport is frequently poor in less-urbanised areas (due to its high costs). However, regional towns (H_2) have higher bicycle use than the largest towns (H_1), which may be explained by the high quality of public transport and the dominance of short commuting trips in H_1 municipalities, which encourages walking. We also suspect that factors such as high volumes of traffic and the risk of bicycle theft deter potential cyclists in large towns.

A classification of municipalities confirms that high proportions of cycling commuters are correlated with low rates of severe/fatal accidents among cyclists. Interestingly, the results revealed a clear-cut north-south division (positive spatial autocorrelation), which hence suggests that different (regional) policies are implemented in terms of bicycle promotion and safety. In Flanders, most municipalities have a high percentage of cyclists and low risks of being seriously injured or killed while cycling to work. The availability of cycle infrastructure, the flat terrain, the high population and job density, as well as the presence of pro-cycling policies may be some of the factors that make this environment quite attractive and safe for cyclists. Cycling is also part of the Flemish lifestyle and cyclists are generally expected and respected by motorists in Flanders. This produces a virtuous circle since better road safety lowers both the actual and perceived risk of cycling and then encourages more cycling, which in turn makes the environment still safer. Moreover, Flemish policy makers invest more in cycle infrastructures, owing to a greater number of cyclists (high demand). This situation in Flanders is similar to that in the Netherlands (Dutch Ministry of Transport, Public Works and Water Management, 2007).

In contrast, the low proportion of commuters cycling to work in Wallonia is often associated with a high risk of accident. Topography, high driving speeds, long commuting distances as well as car-oriented policies and lifestyles are associated with this scenario. Higher accident risks also deter bicycle use: they make the Walloon environment unsafe and consequently unattractive to (potential) cyclists. The lack of high-quality infrastructure as well as the fact that car drivers generally do not expect to see cyclists on the road probably explain the high risks of severe/fatal accident. In addition, motorists may be less respectful towards cyclists, partly because they have never themselves experienced commuter cycling. Each of these background characteristics confirm the fears and the high perceived risk of accident Walloon residents associate with cycling.

Last but not least, inter-municipality differences are observed: risks of severe/fatal accidents for cyclists are higher in less-urbanised environments, while the reverse is true in urban areas. In the latter, the presence of features such as physical barriers (e.g. road humps), congestion, lower speed limits and higher numbers of pedestrians force motorists to slow down and adapt their driving behaviour, which improves the safety of all road users. In particular, it reduces the differential between the speed of fast and slow modes of transport, and so decreases the risk of cyclists suffering from injuries in urban areas. Such urban features hence probably explain why low to moderate values of risk are observed in the Brussels-Capital Region. Interestingly, these latter observations are not in line with the perception of overall danger Brussels' residents have about cycling. It is hence hypothesized here that variables such as the traffic volumes and the lack of high-quality cycleways play a prominent role in deterring potential users from cycling. This latter assumption is further explored (and confirmed) in Chapter 3 of this thesis.

Chapter 3

Spatial determinants of cycle commuting Modelling meso-scale spatial variations in Belgium[8]

Outline

This chapter attempts to explain the spatial variation of the use of a bicycle for commuting to work at the level of the 589 municipalities in Belgium. Regression techniques were used and special attention was paid to autocorrelation, heterogeneity and multicollinearity. Spatial lag models were used to correct for the presence of spatial dependence and a disaggregated modelling strategy was adopted for the northern and southern parts of the country. The results show that much of the inter-municipality variation in bicycle use is related to environmental aspects such as the relief, traffic volumes and cycling accidents. Town size, distance travelled and demographic aspects (e.g. share of youngsters or percentage of households with young children) also have some effect. In addition, there are regional differences in the effects of the structural covariates on bicycle use: the impact of variables such as traffic volume and cycling accidents differs substantially between the north and the south of the country. This chapter also suggests that high rates of bicycle use in one municipality stimulate cycling in neighbouring municipalities, and hence that a mass effect can be initiated, i.e. more cycle commuting encourages even more commuters in the area to cycle. These findings provide some recommendations for decision-makers wishing to promote a shift from car to bicycle use.

[8] This chapter is adapted from the following paper: Vandenbulcke, G., Dujardin, C., Thomas, I., de Geus, B., Degraeuwe, B., Meeusen, R., Int Panis, L. (2011). Cycle commuting in Belgium: Spatial determinants and 're-cycling' strategies. *Transportation Research Part A* 45, 118-137. [http://dx.doi.org/10.1016/j.tra.2010.11.004]

3.1 Introduction

In Belgium, while approximately 21% of commuters live within cycling distance (i.e. less than 5 km) of their work, and 39% make trips of less than 10 km, only 6% of all commuting trips are carried out with a bicycle as the main mode of transport (Verhetsel et al., 2007). The percentage of people who live within 5 km of their work who commute by bicycle is relatively low (19%), and the majority (more than 53%) use their car (Figure 3.1). There is hence great potential for a shift from car to bicycle for short commutes. As suggested in Chapter 2, there are however several societal, economic and environmental factors that dissuade people from cycling and make the environment unattractive and unsafe for cyclists. These include a lack of (appropriate) cycle infrastructure, the topography, weather, road accidents, and company-related constraints (e.g. the need to carry bulk goods and/or to be well-groomed, or the accessibility of the company to public transport). They need to be clearly identified to help policy makers to mitigate them and to promote bicycle use in Belgium. Such findings could then support the implementation of adequate policies in favour of a modal shift from car to bicycle commuting, at least for short distances.

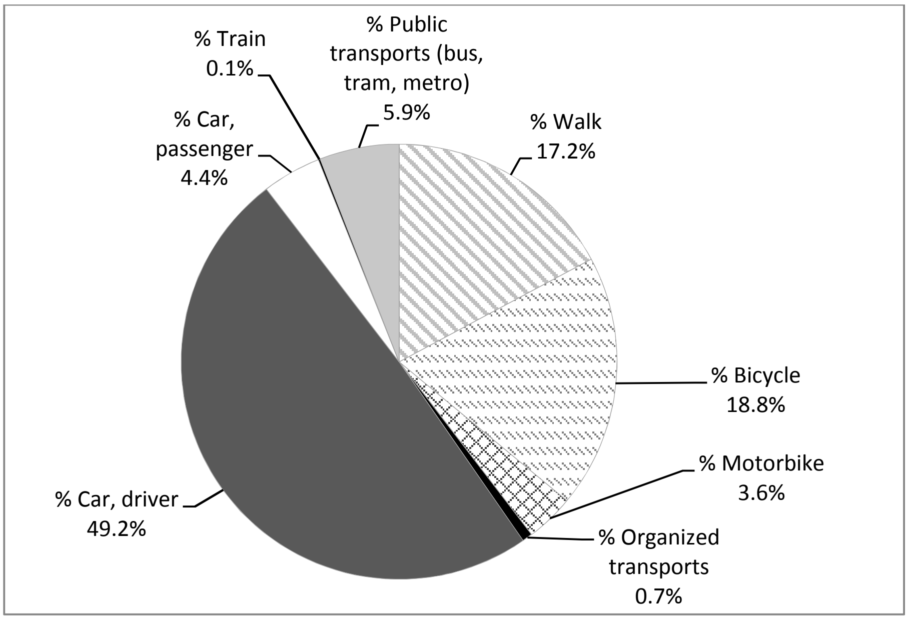


Figure 3.1: Commuters' modal share for distances up to 5 km (Belgium)

This chapter then extends the exploratory spatial data analyses conducted previously on the spatial patterns of bicycle use for commuting (Chapter 2), in order to obtain thorough results as well as statistically-based recommendations for planners and policy makers. Within this framework, we here aimed at examining which factors have the greatest influence on bicycle use for commuting in Belgium. We therefore carried out multivariate analyses at the scale of all 589 municipalities (the smallest administrative unit) in the country. A large set of 'explanatory' variables was included in the analysis, with specific attention to environmental variables as well as demographic components. Spatial autocorrelation, heterogeneity and multicollinearity problems were diagnosed and treated, with the aim of improving the results.

The structure of the chapter is as follows. An exhaustive review of the literature on the factors that have a potential impact on bicycle use is given in Section 3.2. Section 3.3 describes the objectives of the chapter and the data (dependent variable and explanatory variables) in more detail. The methodological approach used to deal with multicollinearity, heterogeneity and spatial autocorrelation is presented in Section 3.4. The results of the multivariate analyses are reported in Section 3.5. In Section 3.6, our concluding remarks underscore the importance of accounting for multicollinearity, spatial dependence and spatial heterogeneity to achieve reliable statistical inferences. Results obtained within the framework of this chapter will feed the formulation of policy recommendations in the general conclusion of this thesis (Chapter 6). Such recommendations overall suggest pro-cycling strategies aiming at increasing bicycle use and making it safer while commuting.

3.2 Identifying the main determinants of bicycle use

A large range of factors have an impact on bicycle use in commuting trips: demographic, socio-economic, cultural, societal, but also environmental and policy-related determinants either act as deterrents or encourage cycling. Based on a large – but not exhaustive – review of the literature, this section provides a short overview of these determinants.

3.2.1 Demographic and socio-economic determinants

Socio-economic and demographic determinants include age, income, gender, education, professional field and status, and family commitments (e.g. having young children). Young commuters (< 25 years) generally have low/medium income and often cannot afford a car, which has a clear impact on their choice.

Moreover, some of them do not have a driving license and have to use public transport or non-motorised forms of transport when they travel to work. The physical abilities of individuals also depend on their age: young commuters are more likely to enjoy good physical health and to cycle more. Gender has an influence on the decision on whether or not to cycle: on average, men cycle to work more often than women, although women travel shorter distances than men (Ortúzar et al., 2000; Dickinson et al., 2003; Heinen et al., 2010). Among other factors, women tend to mention their personal security as a reason for not using a bicycle, and often make more complex trips than men due to family commitments (Pooley and Turnbull, 2000; Dickinson et al., 2003; Rietveld and Daniel, 2004; Gatersleben and Appleton, 2007).

Education also has a strong influence on bicycle use, but this depends on the area being studied. In North America a high level of education is positively associated with cycling (Noël, 2003; Plaut, 2005; Zahran et al., 2008), whereas the opposite effect is observed in Santiago (Chile) (Ortúzar et al., 2000) and Belgium (Hubert and Toint, 2002). Lastly, the professional field and status play a role (Toint et al., 2001; Titheridge and Hall, 2006; Parkin et al., 2008; Heinen et al., 2009; Heinen et al., 2010). For instance, Pucher et al. (1999) showed that in San Francisco lots of messengers are immersed in a cycling culture and use their bicycles in spite of the hilly topography. Bicycle use for commuting is generally high in academic towns (Martens, 2004; Rodríguez and Joo, 2004).

Note finally that other determinants are reported in the literature as influencing bicycle use, although it is to a lesser extent compared to the previous ones. These refer to the marital status (e.g. single, married, widowed), the home characteristics (e.g. parking facilities, garden) and the neighbourhood characteristics (e.g. easily accessible shopping, parks, sport grounds) (for further details, see Meurs and Haaijer, 2001; Pikora et al., 2003; Moudon et al., 2005).

3.2.2 Cultural and societal determinants

The literature often mentions that societal and cultural factors influence bicycle use (see e.g. Jensen, 1999; Pucher et al., 1999; Ortúzar et al., 2000; Rietveld, 2001; Dickinson et al., 2003; Rietveld and Daniel, 2004; Plaut, 2005; Pucher and Buehler, 2006; Zahran et al., 2008; Heinen et al., 2010). A low societal status often tends to be associated with commuter cycling, especially in countries where the car is dominant (e.g. US); utilitarian cycling is often considered as a fringe activity and suffers from a renegade image (Pucher et al., 1999; Moudon et al., 2005). However, the cycling culture is quite developed in some Northern countries of Europe (e.g. the Netherlands and Denmark). Such differences between countries, regions or even ethnicities are probably explained by tradition and lifestyle. A meaningful example is provided by Rietveld and Daniel (2004), who show that immigrants with a different cultural background are unlikely to cycle in the Netherlands and prefer to use public transport or a car. The fact that cycling does not play a prominent role in the native country probably explains such a result as immigrants are probably not yet adapted to the use of the bicycle (from a behavioural point of view) and/or have a different overall perception of cycling (e.g. they may associate it to a low societal status, or to high risks of being injured in a road accident).

3.2.3 Environmental determinants

The main environmental determinants influencing bicycle use are relief, weather (and climatic conditions), urban spatial structure, and infrastructure. Hills influence negatively the attractiveness of non-motorised modes of transport (Noël, 2003; Rodríguez and Joo, 2004; Heinen et al., 2010). Cycling up hills is uncomfortable, requires substantial physical effort (Rietveld, 2001; Gatersleben and Appleton, 2007) and affects travel time in the generalised cost function since it is slower than going down hill or on the flat.

Weather (short-term) and climatic (long-term) conditions are often mentioned in the literature. Low or high temperatures (e.g. extreme heat combined with air pollution), frequent rain, snow, ice and strong winds may act as deterrents to commuter cycling (Nankervis, 1999; Richardson, 2000; Bergström and Magnusson, 2003; Parkin et al., 2008; Zahran et al., 2008; Koetse and Rietveld, 2009; Heinen et al., 2010). Like topography, these factors decrease the level of comfort of cycling and increase the physical effort required.

The urban structure influences the likelihood of commuter cycling through several factors, such as population and job densities, mixed land-use and town size (Kitamura et al., 1997; Rietveld, 2001; Heinen et al., 2010; Verhetsel and Vanelslander, 2010). In urban areas, a high degree of connectivity (i.e. the ability to travel directly), associated with short distances (due to compactness and the presence of mixed-use activities) encourage cycling and walking in commuting trips (Saelens et al., 2003). Distance is an important barrier that limits cycling: only people living close to their workplaces will be interested in cycling (Kingham et al., 2001; Dickinson et al., 2003; Krizek et al., 2010; Verhetsel and Vanelslander, 2010). As suggested by Chapter 2, town size also seems to play a key role: few large towns (with more than 2 million inhabitants) have a bicycle commuting rate exceeding 10%. Medium-sized and compact towns perform better since they contain fewer barriers (e.g. motorways) and traffic densities are lower (Pucher et al., 1999). In the largest towns, proximity to the nearest stop and the high frequency of services however make public transport attractive and highly competitive for distances between 1 and 7.5 km. For instance, the distance to the nearest public transport stop in Brussels (bus, tramway, underground) does not exceed 250m for more than 63% inhabitants, which partly explains the low share of cyclists observed here (Vandenbulcke et al., 2007). Up to 1 km, walking competes strongly with cycling (Pucher et al., 1999; Ortúzar et al., 2000; Witlox and Tindemans, 2004). As mentioned in Chapter 1, the lack of room to store a bicycle in densely occupied buildings may also be another reason to observe low shares of cycling in large towns.

Infrastructure (e.g. cycle lanes and racks) is an essential ingredient for improving bicycle use and cyclists' safety (Hopkinson and Wardman, 1996; McClintock and Cleary, 1996; Rietveld, 2001; Reynolds et al., 2009; Heinen et al., 2010). Well-planned and well-kept infrastructure (through design, maintenance and adequate connectivity) encourages cycling and reduces road accidents and their costs (Aertsens et al., 2010). Depending on the type of planning, several benefits can be provided for cyclists: e.g. improved comfort, reduced travel time, more enjoyment and increased safety. Dedicated paths (e.g. residential streets) as an alternative to main urban roads are an efficient way of reducing the exposure of cyclists to exhaust fumes (Hertel et al., 2008; Thai et al., 2008; Int Panis et al., 2010). Increased safety can also be achieved by developing continuous and designated cycle lanes, and ensuring that cyclists are still visible to motorists; this is often more highly valued by cyclists than other factors (e.g. reduced travel time, easy parking) (Hopkinson and Wardman, 1996; Tilahun et al., 2007). A well-developed network of cycle facilities combined with the provision of bicycle parking facilities at stations/stops may improve the accessibility of public transport to cyclists, and hence, provide a competitive alternative to the car for commuting trips (Martens, 2004; 2007). Finally, the presence of facilities such as covered/secure cycle parking, lockers, showers and changing facilities at the workplace stimulates commuter cycling (Rietveld, 2000; Kingham et al., 2001; Dickinson et al., 2003; Pucher and Buehler, 2006; Van Malderen et al., 2009; Vanoutrive et al., 2009; 2010). Combined with the provision of continuous cycle facilities and a mileage allowance for cycling to work, such facilities are expected to have a significant impact on commuting by bicycle (Wardman et al., 2007; Van Malderen et al., 2009).

3.2.4 Policy-related determinants

Policy-related variables (i.e. planning and pro-cycling policies) play a key role in encouraging more and safer cycling through the implementation of a wide range of measures (Pucher et al., 1999; Rietveld, 2001; Dickinson et al., 2003; Pikora et al., 2003; Pucher and Buehler, 2008; Heinen et al., 2010). Land-use planning can prevent urban sprawl by favouring compact and mixed-use solutions which reduce travelling distances and – consequently – favour the use of non-motorised transport for commuting (Cervero and Kockelman, 1997; Kitamura et al., 1997; Meurs and Haaijer, 2001; Noël, 2003; Titheridge and Hall, 2006; Chapman, 2007; Woodcock et al., 2007; Verhetsel and Vanelslander, 2010). Moreover, transport planning can modify the design and lay-out of transport networks to improve the connectivity of bikeable roads between different destinations. It can increase the directness of travel through the creation of special intersection modifications for cyclists (e.g. by providing priority signalling or advanced stop zones), the suppression of barriers (e.g. foot and cycle bridges over motorways or waterways), the creation of detours for car drivers, and the introduction of traffic-calming or car-free zones in urban centres (Meurs and Haaijer, 2001; Rietveld, 2001; Saelens et al., 2003; Pucher and Dijkstra, 2003; Pucher and Buehler, 2006; 2008). This makes cycling safer by reducing the risk of collision with motorised traffic, but also more convenient by allowing cyclists to avoid detours and traffic jams (Rietveld, 2001). Improving safety is of prime importance as it is well-known that the (perceived) risk of death and injury in traffic crashes strongly discourages people from cycling (Hopkinson and Wardman, 1996; McClintock and Cleary, 1996; Curtis and Headicar, 1997; Jacobsen, 2003; Pikora et al., 2003; Pucher and Dijkstra, 2003; Pucher and Buehler, 2006; Parkin et al., 2007).

The provision of secure facilities (e.g. guarded cycle racks) along with police surveillance are also efficient means of reducing the risk of bicycle theft or vandalism (which are strong deterrents to cycling). A maximum walking distance with respect to busy areas (e.g. stations) is also recommended for bicycle parking facilities so that these latter are continuously visible for others (Martens, 2007). Such enhanced cycling conditions and resulting shifts from car to bicycle (favoured by pro-cycling land-use and transport planning strategies) not only allow reducing the costs related to the urban sprawl, but also help increasing the economic productivity and development of a specific region or country (Litman, 1994, 1995; Buis, 2000; Burchell et al., 2002; Litman, 2004). Indeed, compact and mixed-use patterns as well as improved transport design and lay-out strategies hold the potential to increase the accessibility to facilities and resources (which in turn reduces the attendant transport costs and externalities due to urban sprawl). More interestingly, investments in non-motorised transport modes are also shown to increase the nearby property values and attract residents and companies that yield some value to the environment and sustainable development (Litman, 1994, 2004). New opportunities for employment can then result from such cycling investments and can in turn increase the performance of the local/regional labour market, owing to e.g. the establishment of new companies attracted by the pleasant environment, the growing demand for facilities resulting from the new residents, and the implementation of the new cycling facilities themselves.

Financial measures can also promote non-motorised modes of transport and regulate the use of the private car. The provision of monetary incentives such as a mileage allowance or an employer-paid discount on the purchase of a new bicycle may stimulate the practise of commuting by bicycle (Kingham et al., 2001; van Wee and Nijland, 2007; Wardman et al., 2007). For instance, Wardman et al. (2007) showed that a payment of £2 per day could double the level of cycling in Great Britain. Higher parking fees, reduced space for car users (with increased 'shared space'), fiscal incentives for less polluting cars, higher fuel prices and the implementation of urban tolls (as in London and Stockholm) are some examples of push-measures which can decrease the attractiveness of private car use and encourage a shift to alternative modes of transport, especially when combined with pull-measures (e.g. high-quality cycle facilities, secure parking, …) (Verhetsel, 1998; De Borger et al., 2001; van Wee and Nijland, 2007).

Company-related factors can also encourage or discourage commuting by bicycle, especially through their organisational aspects (e.g. a strong dress code, the need to carry bulky goods, flexible work schedules), location policies, and the availability of facilities (e.g. changing rooms and cycle lockers at the workplace) (Curtis and Headicar, 1997; Dickinson et al., 2003; Heinen et al., 2009; Vanoutrive et al., 2009, 2010). In particular, a remote location, far from any town or public transport, will result in great dependence on the car and will discourage employees from using any other mode of transport (Vanoutrive et al., 2010). Employees are also unlikely to travel to work by public transport or bicycle if their company provides free cars and fuel. Only reducing the provision of company cars and fuel, combined with other measures (e.g. incentives for cycling and public transport), can induce a shift away from the car towards alternative modes of transport (Kingham et al., 2001).

Finally, the promotion of cycling is important, since attitudes towards mobility, the environment, etc., are closely linked to travel behaviour (Kitamura et al., 1997). Such promotion can increase cycling and can be achieved through educational programmes (e.g. teaching cycling safety at schools), promotional events, the active involvement of advocacy groups and town officials (e.g. police officers on bicycles), and up-to-date information for cyclists (e.g. cycling maps showing 'bikeable' roads) (Curtis and Headicar, 1997; Pucher et al., 1999; Pucher and Buehler, 2006; Zahran et al., 2008). In particular, promotional events can create a mass effect providing cyclists with confidence and enthusiasm (Pucher el al., 1999). Linking cycling to health can also be an efficient way of encouraging more commuters to cycle, since regular exercise improves fitness and health (de Geus et al., 2008a; 2009). Cycling is indeed a low-cost way to tackle health problems linked to physical inactivity (e.g. diabetes, cardio-vascular diseases and cancers). It has also been shown to improve mental health and productivity at work (Pucher et al., 1999; EC, 2000; van Wee and Nijland, 2007).

3.3 Objectives and data

The main aim of this chapter is to explain the variation of the proportion of commuters who travel by bicycle (dependent variable, y), as measured at the scale of the 589 municipalities in Belgium[9]. Explanatory variables used in the multivariate analyses fall into three main categories (demographic and socio-economic, policy-related, and environmental) and refer to most of the determinants identified in Section 3.2. Appendix B.1 lists and describes the explanatory variables. Most of the demographic and socio-economic variables either come from the 2001 census or are obtained from the website of the Directorate-General Statistics and Economic Information (DGSEI). The 2001 census is a self-administered questionnaire, carried out by the DGSEI (DGSEI, 2001b; DGSEI, 2004). It is preferred to other surveys since it is the most recent database and covers the entire population. Environmental and policy-related variables come from a wide range of sources. These latter not only result from policy decisions (e.g. land-use and transport-related measures), but also characterise the 'environment' in which commuters live and travel. Some of these variables (e.g. population and job densities, average commuting distance, town size, the percentages of urban/forest/agricultural land, etc.) are proxies for the urban structure, land use and accessibility of activities/facilities in the municipality. Others (such as the risk of accidents to cyclists, traffic volumes, dissatisfaction with cycle facilities, hilliness, and air pollution) are representative of the overall convenience of cycling in the municipality. Note that further information on these variables is provided in the following subsections.

3.3.1 Demographic and socio-economic variables

Most of the demographic and socio-economic variables come from the 2001 census, carried out by the DGSEI. This census was of great help to compute the

[9] Note that y is continuous, non-negative and constrained to a specific range. Linear models could be hence less suited here. Satisfactory results are however obtained and suggest that the methodological approach adopted here is suitable.

percentage of working people belonging to various life phases (e.g. being less than 25 years old or between 45 and 54) or having specific education levels (e.g. percentage of working people having a university degree as their highest qualification). Interestingly, these data show that a large majority of commuter cyclists have a secondary (60%) or primary (7%) education as highest degree in Belgium. They also indicate that the proportion of commuting by bicycle (9.1%) is the highest within the population of working people having a primary degree as their highest qualification (university degree: 6.4%). Such observations differ from the results obtained by Plaut (2005), who shows that a higher education encourages cycling in the USA.

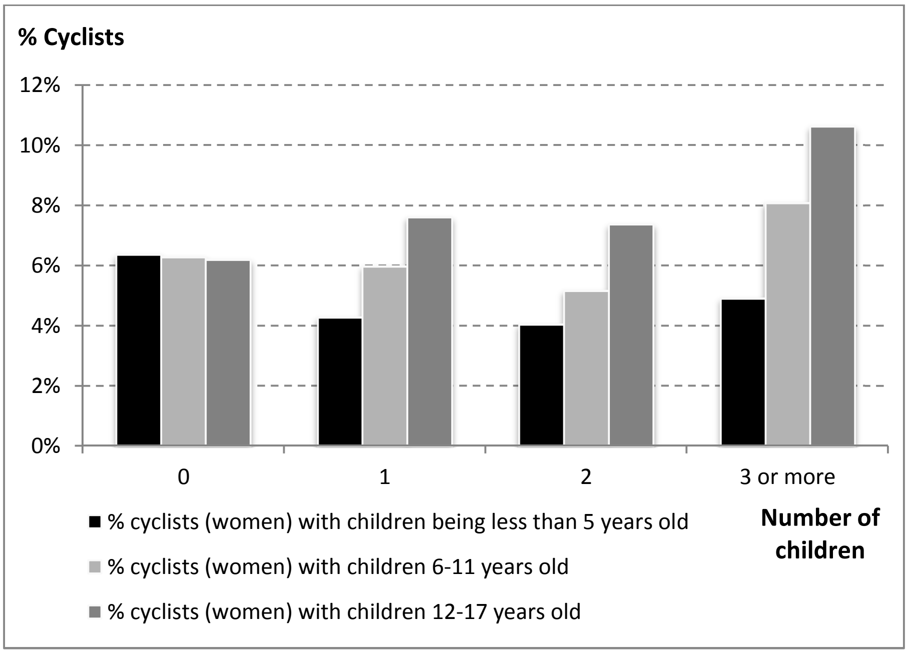


Figure 3.2: Percentage of commuter cyclists (women) having children being less than 5 years old, 6-11 years old, or 12-17 years old

Another variable extracted from the 2001 census is related to the subjective health of people, and consequently with the (physical and/or mental) ability to cycle. This variable is the percentage of inhabitants in a municipality feeling they have a bad state of health. The census also provides the opportunity to get a proxy for family commitments: the percentage of working households (i.e. with one or more working parents) having one or more young children (i.e. being less than 5 years old). Exploratory analyses suggest that the presence of young children in the household discourages cycling, especially for (working) women.

For example, Figure 3.2 exhibits that the deterrent impact is the highest when women have 2 young children, whereas it is the lowest when they have 3 young children (or more). For households with children being between 12 and 17 years old, such a deterrent effect is however not observed.

Finally, the website of the Directorate-General Statistics and Economic Information (DGSEI) is also used to extract the following variables: median income, percentage of working people who are men (proxy for gender), and percentage of households that do not own any car.

3.3.2 Environmental and policy-related factors

Variables characterizing the 'overall' environment of the municipality are here considered. Population and jobs densities, town size, as well as proportions of urban, forest, agricultural, public and recreational land surfaces by municipality are extracted from the DGSEI website as a first subset of environmental and policy-related variables. In particular, the presence of public facilities is expected to stimulate cycling. Exploratory analyses conducted outside the framework of this chapter (not illustrated here) suggest that most cyclists work in the public field (e.g. administration, education and health). According to Wendel-Vos et al. (2004), it is also assumed that high proportions of forest or recreational areas encourage cycling. Regarding the proxy for the town size, ranks are attributed to the municipalities on the basis of an index and based on the degree of equipment of the municipality as well as on its attractiveness (Van Hecke, 1998). As in chapter 2, this variable is coded in such a way that the largest towns have low values (Brussels is ranked 1; the ranks 6, 7 and 8 are attributed to the smallest and least-populated municipalities).

Accessibility/separation variables are considered as a second subset of data and include the minimum network distance to the closest town (km) as well as the percentage of commuters that live no further than 10 km from their workplace. Commuting distance is also computed from the 2001 census as the observed average distance between residence and workplace (by municipality of residence) since it highly constrains the transport mode choice (Kingham et al., 2001; Dickinson et al., 2003; Saelens et al., 2003). It is here assumed that commuters are more likely to cycle when they live in urban environments with moderate and high densities. Municipalities with low densities are characterised by large commuting distances and are hence not attractive for cycling. However, municipalities with high densities are not necessarily associated with high bicycle shares because of the short distances between activities (which favours walking trips) and the high-quality of public transport (Ortúzar et al., 2000). Housing characteristics may also play a role here in the sense that flats have generally smaller floor areas in the densest municipalities. Room is hence often lacking to store bicycles and may then preclude from cycling (see Chapter 1).

Variables referring to the overall convenience of cycling in the municipality make up the last subset of environmental and policy-related variables. A proxy summarizing the quality of cycle facilities in the municipality – i.e. the percentage of households estimating they have low-quality cycle facilities located in their neighbourhood – is first extracted from the 2001 census. More concretely, it appraises how unsatisfied households are about the neighbouring cycle facilities observed in the municipality. Secondly, the risk for a cyclist of being involved in a road accident was also roughly estimated using the 2001 census and DGSEI data. This risk is defined in a same way as in chapter 2: $R_i = N_i / T_i$, where N_i is the average annual number of (all) accidents for cyclists aged between 18 and 65 years (occurring between 2002 and 2005 and on weekdays in municipality i) and T_i is the total time spent travelling by commuter cyclists living in municipality i per year. Thirdly, data on bicycle theft (2000-2002) are also obtained from the Federal Police. In Belgium, criminality statistics show that approximately 32,000 bicycle thefts occur each year (although, in reality, only 45% victims lodge a complaint). A ratio between the number of thefts and the number of cyclists is then computed with the aim to estimate the risk of bicycle theft by municipality. Fourthly, traffic data are obtained from counting, surveys and estimations carried out by the Federal Public Service (FPS) Mobility and Transports[10]. Such data are used to compute a proxy for the volume of traffic transiting in a municipality, which is expected to have a deterrent impact on bicycle use when the traffic volume is high. This proxy is here expressed as the number of vehicles-km by kilometre of municipal or regional road (motorways are excluded since they are not 'bikeable'). Fifthly, the ambient air quality is also taken into account in order to examine the relationship between concentrations of air pollutants and bicycle use for commuting trips. Particulate matter concentrations (PM10) are obtained from the Belgian Interregional Cell for the Environment (IRCEL-CELINE) for the years 2000-2005. Measurements are made in telemetric stations and interpolated to a grid data formed by pixels of 4 x 4 km. Performing areal statistics from grid data then allows to estimate the mean concentration of particulate matter (PM10) by municipality. Last but not least, the mean slope along the ('bikeable') road network is computed for each municipality, using a Digital Elevation Model (DEM) collected from the National Geographic Institute (NGI). Such a DEM corresponds to a set of height values assigned to pixels (90 x 90m) and is incorporated into a Geographic Information

[10] Estimations are based on the size of the automobile park and on the volume of traffic transiting on the neighbouring road sections.

System (GIS) in order to compute slopes. In ArcGIS 9.2 (tool 'Slope'), these latter are defined as the maximum rate of change from each cell (or pixel) to the closest neighbours. The final step then consists in estimating the mean slope along the municipal and regional road networks (from which motorways and express roads are excluded since they are not allowed for bicycle traffic).

3.3.3 Limitations

Appendix B.1 is far from being exhaustive. In particular, the societal and cultural variables described in Section 3.2 were not included in this analysis, except through the integration of spatial regimes in the final model[11]. It is also noteworthy that data on immigrant background (or ethnic origin) are quite tricky to use here as proxies for the travel behaviour, since many non-native residents obtained the Belgian nationality (without subsequently changing their travel habits) (Deboosere et al., 2009). Finally, weather- and/or climatic-related variables (e.g. wind, rainfall, temperature) are not included in the model since the quality of the data is (spatially) poor (i.e. data are collected over a limited number of measurement stations). Also, in small countries such as Belgium, it is to be expected that few spatial variation exists between the municipalities. As illustration, insignificant estimates were obtained within the framework of a study conducted at the scale of the Dutch municipalities (Rietveld and Daniel, 2004). Collinearity with topography is also expected to occur, which suggests that incorporating such weather-/climatic-related variables in the model would not yield more explanatory power.

3.4 Methodology

A combination of exploratory (spatial) data analyses and spatial econometric techniques is here considered, taking advantage of the use of several specialized software packages (SAS, GeoDa and R). Descriptive statistics and bivariate analyses (i.e. Pearson and Spearman's rank correlation) are first computed in order to explore the relationships between each of the explanatory variables and the dependent variable y. Multivariate models are then applied, with the aim of examining the relative importance of the explanatory variables for the spatial variation in bicycle use (at the scale of municipalities). These are described in the following subsections. To improve the statistical inference process, special

[11] Exploratory spatial data analyses (ESDA) do indeed suggest that the regimes/clusters defined in Section 3.5.5.1 are representative of different cultures (the Flemish-Walloon split).

attention is paid to multicollinearity, spatial heterogeneity and spatial autocorrelation.

3.4.1 Ordinary least squares model

For N observations and K exogenous independent variables, the structure of the first model (OLS) in matrix form is as follows:

$$y = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3.1}$$

where y is a $N \times 1$ vector of observations i on the dependent variable (proportion of commuting by bicycle in municipality i), $\boldsymbol{\beta}$ is a $K \times 1$ vector of coefficients for the independent variables, X is a $N \times K$ matrix of observations i on the independent variables (including a constant term), and $\boldsymbol{\varepsilon}$ is a $N \times 1$ vector of error terms at location i. In this chapter, N is equal to 589 (the number of municipalities in Belgium).

The first step in testing the validity of the ordinary least squares (OLS) model was to compute condition indices, tolerance and variance inflation factor (VIF) values so as to diagnose the existence of multicollinearity. The major assumptions of the regression (linearity, homoscedasticity, normality and spatial independence of the residuals) were then tested. The White, Breusch-Pagan and Koenker-Bassett tests for the presence of non-constant variance in the errors (heteroskedasticity) were performed first, and the asymptotic version HC3 of the heteroskedasticity-consistent covariance matrix (HCCM) was then used to correct for heteroskedasticity (Long and Ervin, 2000). HC3 was preferred over HC0, HC1 and HC2 owing to its better properties for testing estimates that are most affected by heteroskedasticity (Long and Ervin, 2000). Note that this correction is most commonly known as the White's correction.

3.4.2 Spatial autoregressive modelling

Spatial autocorrelation is another misspecification affecting the results of the OLS regression since it will lead to a wrong statistical interpretation (e.g. biased and inconsistent coefficients, biased t- and F-statistics, misleading measures of fit) (Anselin, 1992). This implies that a functional relationship exists between a municipality i and the neighbourhood, i.e. the value (e.g. residuals or observations of the dependent variable) in a municipality i depends on the values observed in the 'neighbouring' municipalities. Spatial autoregressive modelling (SAR) was then used to deal with the presence of spatial autocorrelation. It is divided into two alternative specifications: spatial error and spatial lag models.

While the first specification suggests the presence of omitted explanatory variables, the second indicates the possibility of a diffusion process (i.e. an event in one municipality increases the likelihood of the same event occurring in neighbouring municipalities).

The spatial error model (SEM) specifies a spatial autoregressive process for the error term ε to account for the spatial influence of unmeasured (or omitted) explanatory variables on the proportion of commuting by bicycle in neighbouring municipalities. In matrix form, it is formally expressed as:

$$y = X\beta + \varepsilon \tag{3.2}$$

$$\text{with } \varepsilon = \lambda W\varepsilon + \xi \tag{3.3}$$

where λ is a spatial autoregressive coefficient, W is a $N \times N$ spatial weights matrix (row-standardised) and ξ is a white noise error. By contrast, the spatial lag model (SLM) assumes that the dependent variable in municipality i is influenced by the values of the dependent and independent variables in the surrounding municipalities j. The magnitude of this spatial influence (or 'spillover effect') is captured by a spatial autoregressive coefficient ρ. In matrix notation, the SLM specification is:

$$y = \rho Wy + X\beta + \varepsilon \tag{3.4}$$

where Wy is the spatially lagged endogenous variable. Interestingly, independent variables for which coefficient values decrease (in absolute terms) after including ρ are expected to have some influence on y_i from the neighbouring municipalities j of i (in addition to having an in-municipality effect).

Note that for both specifications (SLM and SEM), the standard R^2 is invalid since a maximum likelihood (ML) estimation is used. Some more appropriate measures of fit are the log-likelihood, the Akaike information criterion (*AIC*) and the Schwarz information criterion (*SIC*) (Anselin, 1988; Anselin, 2005).

3.4.3 Diagnostics for spatial autocorrelation

Spatial dependence in OLS specifications can be detected using either the Moran's I statistic or the Lagrange Multipliers (*LM*). Moran's I statistic is the most commonly used measure to detect for the presence of spatial error dependence in OLS regression (Moran, 1948; Anselin, 1988). It takes the following form:

$$I = (N/S_0)(e'We/e'e) \tag{3.5}$$

where S_0 is the sum of all weights and e is a $N \times 1$ vector of residuals. This statistic is however inappropriate in the presence of heteroskedastic or non-normally distributed errors[12], but also in suggesting which specification (SEM or SLM) should be used to correct for spatial autocorrelation (Anselin and Rey, 1991; Anselin, 2005). Lagrange Multiplier (LM) diagnostics and their robust forms (Robust LM) are then performed instead of the Moran's I, especially because they help to identify the form of spatial dependence (spatial error or spatial lag). The LM diagnostics for lag (LM_{lag}) and error dependence (LM_{error}) are expressed as:

$$LM_{lag} = \left[Ne'W_1y/e'e\right]^2 \left[N(W_1Xb)'M(W_1Xb)/e'e + tr\left(W_1'W_1 + W_1^2\right)\right]^{-1} \sim \chi^2(1) \quad (3.6)$$

$$LM_{error} = \left[Ne'W_2e/e'e\right]^2 \left[tr\left(W_2'W_2 + W_2^2\right)\right]^{-1} \sim \chi^2(1) \quad (3.7)$$

where $M = I - X(X'X)^{-1}X'$, b is a vector of OLS estimates of β, tr is the matrix trace operator, W_1 is the spatial weights matrix for the spatially lagged dependent variable and W_2 is the spatial weights matrix for the spatially lagged error term (Anselin and Rey, 1991; Anselin and Florax, 1995). The robust LM diagnostics are also advised to be estimated since they are robust to the presence of spatial lag (resp. error) when diagnosing for spatial error (resp. lag) dependence (Anselin et al., 1996). An analysis of the two kinds of diagnostics (robust and non-robust) finally determines which spatial autoregressive model is the most convenient to deal with spatial autocorrelation (Anselin and Florax, 1995).

Note that both Moran's I statistic and LM test are based on the assumption that the error terms follow a normal distribution. In empirical works, the Jarque-Bera statistic is commonly used to test this assumption of normality. If the errors turn out to be non-normally distributed, a useful alternative is the Kelejian-Robinson diagnostic (Kelejian and Robinson, 1992).

3.4.4 Diagnostics for heteroskedasticity in presence of spatial autocorrelation

Spatial dependence invalidates the distributional properties of several parametric tests for heteroskedasticity. In particular, the power and the empirical rejection frequencies for the White and Breusch-Pagan tests are strongly affected when the error terms are spatially correlated (Anselin, 1988). As a consequence, two

[12] The presence of non-normally distributed errors leads to an under-rejection of the null hypothesis of the Moran's I diagnostic, whereas the presence of heteroskedasticity results in an over-rejection of the null.

alternative strategies are proposed to test for the presence of heteroskedasticity: the joint *LM* test and the Spatial Breusch-Pagan test. In the presence of heteroskedasticity and spatial autocorrelation, the joint *LM* test is more powerful than the individual statistics used to test for both spatial effects. It diagnoses for the joint presence of heteroskedasticity and spatial autocorrelation and is obtained by summing a Breusch-Pagan statistic and an *LM* test against residual autocorrelation. If the joint null hypothesis of the test is rejected, the individual tests could be separately performed in order to identify the origin of the rejection (Anselin, 1988; Anselin and Griffith, 1988). Finally, the Spatial Breusch-Pagan test is a spatially adjusted version of the Breusch-Pagan test and consists of carrying out a test for heteroskedasticity while accounting for the presence of spatial dependence. Further details about these tests are provided by Anselin (1988) and Le Gallo (2004).

3.4.5 Spatial heterogeneity and regimes

Spatial heterogeneity was taken into account in a number of ways. These include focusing on the issue of heteroskedasticity (see Sections 3.4.1 and 3.4.4), and testing for the structural stability of coefficients between spatial subsets of the data (spatial regimes). In the presence of structural instability, the parameter estimates take on different values in distinct geographic areas. Formally, a regression with Regimes 1 and 2 (e.g. north and south) is called a spatial regime regression. It is expressed as:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \tag{3.8}$$

where y_1 and y_2 are the vectors of observations of the dependent variables, β_1 and β_2 are the vectors of coefficients of the independent variables, X_1 and X_2 are the matrices of observations of the independent variables (including a constant term for each regime), and ε_1 and ε_2 are the vectors of error terms for Regimes 1 and 2 respectively. If spatial dependence persists after the spatial heterogeneity has been modelled, the spatial regime specification should also account for spatial autocorrelation. Equation (3.4) hence takes on the form:

$$\text{Regime 1:} \qquad y_1 = \rho W_1 y_1 + X_1 \beta_1 + \varepsilon_1 \tag{3.9}$$

$$\text{Regime 2:} \qquad y_2 = \rho W_2 y_2 + X_2 \beta_2 + \varepsilon_2 \tag{3.10}$$

where W_1 and W_2 are the spatial weights matrices for Regimes 1 and 2, respectively (Le Gallo, 2004; Anselin, 2007; Bivand, 2008).

3.4.6 Diagnostics for structural instability

The stability of the coefficients across regimes can be diagnosed using the Chow test. The null hypothesis of this test is based on the constraint that the coefficients do not vary across regimes, i.e. there is a regional homogeneity (H_0: $\beta_1 = \beta_2$). The test statistic C is expressed as:

$$C = \left[\left(e_R' e_R - e_U' e_U \right) / K \right] / \left[e_U' e_U / (M - 2K) \right] \sim F(K,\ M - 2K) \tag{3.11}$$

where K is the number of regressors, M is the number of regimes, and e_R and e_U are the OLS residuals from a restricted model (where $\beta_1 = \beta_2$) and from an unrestricted model respectively. This test is however invalid when the error terms are spatially autocorrelated and must consequently be corrected using asymptotic procedures. This yields an asymptotic spatially adjusted test for structural stability, also called the Spatial Chow test (C_G). When the errors terms follow a spatial autoregressive process, it takes on the following form:

$$C_G = \left[e_R' (I - \lambda W)' (I - \lambda W) e_R - e_U' (I - \lambda W)' (I - \lambda W) e_U \right] / \sigma^2 \sim \chi^2(K) \tag{3.12}$$

where I is the identity matrix, λ is the ML estimate for the spatial autoregressive parameter, and σ^2 is the estimate for the error variance for either the restricted model, the unrestricted model, or both (Anselin, 1988).

The presence of structural instability can also be detected and visualised using an exploratory spatial data analysis (ESDA). This helps to identify the presence of global and local patterns of spatial autocorrelation and heterogeneity (e.g. spatial outliers or clusters) in the proportion of commuting by bicycle (Anselin, 1998; Le Gallo and Ertur, 2003; Baller et al., 2001; Ramajo et al., 2008). ESDA can be undertaken by performing common measures of spatial autocorrelation, such as the Moran's I statistic, the Moran scatterplot and the local indicators of spatial association (LISA) (Anselin, 1995; Le Gallo, 2004). While the Moran's I is a quite global statistic[13], the Moran scatterplot and LISA may yield more specific insights into the presence of local patterns of spatial autocorrelation and instability.

In particular, the Moran scatterplot plots the spatially lagged variable Wy against the dependent variable y and allows the local spatial association (between a municipality and its neighbours) to be categorised into four groups: HH (municipality with a high value surrounded by municipalities with high values), LH (low value surrounded by high values), LL (low value surrounded by

[13] It only gives a formal indication on the degree of linear association between the dependent variable y and the spatially lagged variable Wy.

low values) and HL (high value surrounded by low values). These groups correspond to the four quadrants of the Moran scatterplot[14]. HH and LL refer to spatial clusters (positive spatial autocorrelation), while LH and HL indicate spatial outliers (negative spatial autocorrelation). Finally, the information derived from the categorisation into four groups, combined with that resulting from the computation of the significance values of LISA yields the LISA cluster map. This gives an indication of the location of significant spatial clustering and diagnoses local instability (e.g. pockets of non-stationarity). It hence facilitates the identification of spatial outliers and spatial regimes (Le Gallo and Ertur, 2003; Baller et al., 2001). Note that a categorisation process into more than four groups/regimes could probably be carried out, although it is not performed here as: (*i*) the number of observations becomes smaller for each regime (which in turn reduces the statistical significance of parameter estimates); (*ii*) the interpretation of the different groups/regimes is made more complex and trickier, as it would involve defining different levels/degrees of spatial clusters and outliers; (*iii*) to our knowledge, there are no theoretical grounds suggesting the use of more than four groups/regimes in a spatial regime model.

3.5 Results and discussion

This section presents and discusses the results of the multivariate analyses, with the aim to explain the spatial variation in the proportion of commuting by bicycle at the scale of the municipalities. As previously mentioned, special attention is paid to multicollinearity, spatial heterogeneity and spatial autocorrelation. Also note that quite similar results are obtained when control is made over the presence of various types of spatial interactions in the model (i.e. direct and indirect), or when using the proportion of cyclists among commuters who travel less than 10 km as dependent variable[15] (see Section 3.5.6). In this latter case, the analysis of the results allows examining whether or not changes in the estimates occur in the case where short distances ($\leq$ 10 km) are considered for commuting.

[14] The quadrants are delimited by the axes $y = 0$ and $Wy = 0$, where y is standardized and W row-standardized.
[15] Such a 10km threshold is selected on the basis of descriptive/exploratory data analyses conducted in Chapter 2. This latter confirmed that, for most people, 10 km is the limit for cycling to work.

Table 3.1: Basic statistics and bivariate correlations with the proportion of commuting by bicycle at the scale of the municipalities ($N = 589$)

Variables	Mean	SD	Min	Max	Correlation with y
Dependent variables					
% cycle commuting (y)†	4.6	4.6	0.0	21.7	-
Independent variables					
% working men	57.6	2.0	50.7	64.6	0.00
% age 1 (< 25)	10.0	1.9	5.2	17.5	0.54***
% age 2 (45-54)	23.5	2.4	15.7	42.4	-0.39***
% age 3 (> 54)†	6.9	1.5	3.9	15.3	-0.30***
% young children (≤ 5 years)	20.7	2.7	10.5	30.6	-0.39***
% education 1 (primary school)	6.0	1.9	2.0	15.3	0.05
% education 2 (secondary school)†	57.5	7.3	25.8	70.3	0.21***
% education 3 (university degree)†	36.6	8.4	15.3	71.8	-0.20***
Income	19.4	2.0	13.4	25.1	0.25***
% bad health	24.1	5.0	15.1	39.4	-0.58***
% car owner†	18.1	6.9	8.1	57.1	-0.25***
Population density†	675.6	1735.7	21.4	19128.6	0.28***
Jobs density†	203.6	725.9	1.3	8342.1	0.38***
Commuting distance	22.6	5.7	10.2	42.7	-0.54***
Town distance†	14.7	12.2	0.0	85.8	-0.26***
% short cycle commuting (≤ 10km)†	35.9	11.3	12.8	67.1	0.46***
Town size (urban rank)[a]	6.3	1.5	1.0	8.0	-0.23***
% urban areas†	28.4	19.5	4.7	99.5	0.34***
% forested areas†	14.3	16.2	0.0	74.0	-0.33***
% agricultural areas	57.3	21.2	0.5	93.6	0.09**
% public services areas†	1.0	1.7	0.0	22.9	0.17***
% recreational areas†	2.0	2.4	0.1	15.9	0.12***
Slope†	2.8	2.0	0.7	10.8	-0.77***
% dissatisfaction with cycle facilities	65.1	18.4	24.6	95.8	-0.82***
Bicycle theft†	56.4	166.8	0.0	2451.7	0.75***
Theft risk	8.9	5.8	0.0	33.0	0.05
Accident risk†	0.3	0.5	0.0	7.0	-0.32***
Air pollution	29.3	4.2	20.6	40.8	0.23***
Traffic volume 1 (regional roads)†	3.1	1.9	0.0	14.0	0.31***
Traffic volume 2 (municipal/local roads)†	0.2	0.2	0.0	1.4	0.12***

** Significant at the 95% level; *** Significant at the 99% level
n.s.: not significant at the 90% level
SD: Standard Deviation
†: logarithmically transformed variables
[a] Spearman rank correlation

3.5.1 Basic statistics and bivariate correlations

Table 3.1 presents some basic statistics as well as Pearson and Spearman's rank correlation coefficients between each of the explanatory variables and the dependent variable y (proportion of commuting by bicycle in a municipality i)[16]. Note that y and several explanatory variables are transformed using the logarithmic function $ln(x+1)$ in order to satisfy the assumption of normality[17]. Most variables are significantly correlated with the dependent variable and exhibit the expected signs. The highest correlations are observed for variables related to the dissatisfaction of cycle facilities (–0.82), hilliness (–0.77) and bad health of inhabitants (–0.58). At the scale of the municipalities, these results hence suggest that such variables strongly discourage bicycle use for commuting trips. Interestingly, a positive correlation is also obtained between the dependent variable and the number of bicycle thefts (0.75), which does not highlight the deterrent effect of thefts on cycling. Instead, it indicates that a high number of bicycle thefts is to be found where bicycle use is high (expectable). Other variables show strong relationships with cycling. In particular, the commuting distance (–0.54), the percentage of working households having one or more young children (–0.39), the percentage of working people who are between 45 and 54 years old (–0.39), the percentage of the municipality which is forested (–0.33) and the accident risk (–0.32) are all negatively correlated to the proportion of commuting by bicycle in the municipality. At the opposite, the percentage of working people who are less than 25 years old (0.54), the density of jobs (0.38), the percentage of the municipality which is urbanised (0.34) and the traffic volume on regional roads (0.31) all show positive correlations with the dependent variable. Overall, these relationships confirm hypotheses about mode choice processes in transport geography.

3.5.2 OLS results

A multivariate regression is here applied using OLS estimation (stepwise) and paying special attention to the heteroskedasticity and multicollinearity issues. The analysis of condition indices, tolerance and *VIF* values is helpful to lower multicollinearity as much as possible. The Breusch-Pagan and White tests for heteroskedasticity (Table 3.2) reveal the presence of non-constant error variance in the model; this was corrected using White's correction (HC3). Results for the White-corrected OLS estimation are reported in Table 3.3. They indicate quite

[16] Spearman's rank correlation is only computed for the 'town size' variable (ordinal).

[17] The Pearson's correlation technique assumes that both the dependent and independent variables come from normally distributed populations (Ebdon, 1985).

high goodness-of-fit ($R^2 = 0.879$) and show that most of the parameters are at least significant at a 10% level of probability. The diagnostics do, however, show the presence of spatial dependence (which affects the validity of the OLS estimations). Moran's I statistic and Lagrange Multipliers (LM tests) are indeed highly significant and suggest that spatial autocorrelation is a concern. The joint LM tests also show strong evidence for spatial dependence and confirm the presence of heteroskedasticity. A comparative analysis of the significance of the robust and non-robust forms of the LM tests finally suggests that the spatial lag model is a better way of addressing the spatial autocorrelation issue.

Table 3.2: Regression diagnostics for the OLS and ML estimations

	OLS	ML
Diagnostics for normality		
Jarque-Bera test	4.62	n.a.
Diagnostics for multicollinearity		
Variance inflation value (maximum value)	3.30	n.a.
Condition index (intercept adjusted)	5.03	n.a.
Diagnostics for heteroskedasticity		
Breusch-Pagan test[1]	31.33***	36.87***
Koenker-Bassett test[1]	28.32**	28.82***
White test	213.64***	n.a.
Breusch-Pagan test (north v. south)[1]	88.68***	25.08***
Diagnostics for spatial dependence		
Moran's I of residuals[2]	0.34***	0.01
Lagrange multiplier (lag)	253.37***	n.a.
Robust LM (lag)	86.74***	n.a.
Lagrange multiplier (error)	181.96***	n.a.
Robust LM (error)	15.33***	n.a.
Diagnostics for spatial dependency and heteroskedasticity		
Joint test LM	213.29***	n.a.
Tests on overall stability		
Chow structural instability test[1]	14.20***	120.49***
Diagnostics for residual autocorrelation		
LM test	n.a.	0.00

*Significant at the 90% level; **Significant at the 95% level; ***Significant at the 99% level
n.a.: no test available
[1] Spatial version of the test
[2] Inference computation based on 9999 permutations (for ML estimation only)

Table 3.3: Regression coefficients for the OLS and ML estimations

	OLS, with heterosk. correction	ML, with heterosk. correction
Intercept	6.4124***	3.2698***
	[0.0000]	[0.0000]
Lag coefficient (ρ)	-	0.6015***
	-	[0.5483]
Demographic variables		
% working men	0.0472***	0.01673**
	[0.1150]	[0.0408]
% age 2 (45-54)	-0.0460***	-0.02505***
	[-0.1352]	[-0.0737]
% age 3 (> 54)†	-0.2054*	-0.14503*
	[-0.0456]	[-0.0322]
% young children (≤ 5 years)	-0.0567***	-0.0218***
	[-0.1865]	[-0.0716]
Socio-economic variables		
% education 3 (university degree)†	-0.4988***	-0.23034***
	[-0.1261]	[-0.0582]
Income	0.0030	0.00852
	[0.0072]	[0.0206]
% bad health	-0.0521***	-0.0189***
	[-0.3124]	[-0.1133]
Environmental & policy-related variables		
Commuting distance	-0.0114***	-0.00652**
	[-0.0789]	[-0.0450]
Town size (urban rank)	-0.0954***	-0.0875***
	[-0.1750]	[-0.1604]
Traffic volume 2 (municipal/local)†	-0.9216***	-0.4695***
	[-0.1341]	[-0.0683]
Slope†	-0.4873***	-0.1763***
	[-0.2655]	[-0.0961]
% dissatisfaction with cycle facilities	-0.0127***	-0.0049***
	[-0.2818]	[-0.1077]
Accident risk†	-0.1673**	-0.14495***
	[-0.0500]	[-0.0434]
Air pollution	0.0141***	0.00405
	[0.0717]	[0.0206]
N	589	589
R-squared (R^2)	0.879	-
Adjusted *R*-squared	0.876	-
F-value	297.80***	-
Log likelihood	-102.43	33.68
Akaike information criterion (*AIC*)	236.86	-35.36
Schwarz information criterion (*SIC*)	306.91	34.70

*Significant at the 90% level; **Significant at the 95% level; ***Significant at the 99% level
- : variable not included in the model
Standardised regression coefficients are given in brackets
†: logarithmically transformed variables

3.5.3 Choice of the spatial weight matrix

The computation of a spatial autoregressive model requires the definition of a spatial weight matrix. Here, a 'queen' contiguity-based matrix is used because it provides the best fit and results in a model satisfying to the finite sample condition (Wald test ≥ likelihood ratio ≥ Lagrange multiplier) (Anselin, 1988). In such a case, the elements w_{ij} of the weights matrix W are equal to 1 when the municipalities have common borders and vertices, and 0 otherwise. By convention, the diagonal elements of W take on a zero value since a municipality is defined as being not contiguous to itself. Also, the matrix is row-standardized for ease of interpretation (Anselin, 1988). Note that spatial weight matrices based on the rook contiguity (first order or higher) and Euclidean distances (from 12 to 300 km) were also considered, although they were ruled out since they provided either a lower fit or did not satisfy to the condition for finite samples. Although the queen matrix used here is generally less suited to spatial units characterized by various polygon sizes, it is here thought that it may well reflect the fact that cyclists adopt different travel behaviours depending on the location where they live (e.g. cyclists living in central business districts generally travel short distances owing to the high proximity to facilities, whereas cyclists living in rural areas overall travel larger distances because of the low densities (large separation distances between activities)) (also see Section 5.3.1 for further information). Such a variation in polygon size may indeed be advantageous here as it mainly depends on the type of environment/neighbourhood (urbanized municipalities generally have small polygon sizes, while rural municipalities often have large polygon sizes). Contiguity-based matrices hence probably better mirror the spatially varying cycling behaviour (and, consequently, the spatial relationships) between municipalities. At the opposite, distance-based matrices do not incorporate such a spatially varying effect and may not necessarily be the best choice in our case (which probably partly explains why they were ruled out).

3.5.4 Spatial lag results

The results for the spatial lag models are presented in Table 3.3. White's correction is again used to treat the model for the presence of heteroskedasticity (given that it is detected by the spatial Breusch-Pagan and Koenker-Bassett tests). As the significant Jarque-Bera statistic in Table 3.2 suggests, the ML estimation is valid since the error terms are normally distributed. The same is true for the LM tests and Moran's I statistics. The spatial lag model gives a better fit than OLS (as shown by the log-likelihood that increases from −102 to 34). Moreover, the Moran's I and the LM test statistics both indicate that including a spatially lagged variable in the model eliminates spatial dependence. Figure 3.3 tends to confirm such a result, showing that the residual spatial autocorrelation is strongly reduced in the spatial lag specification.

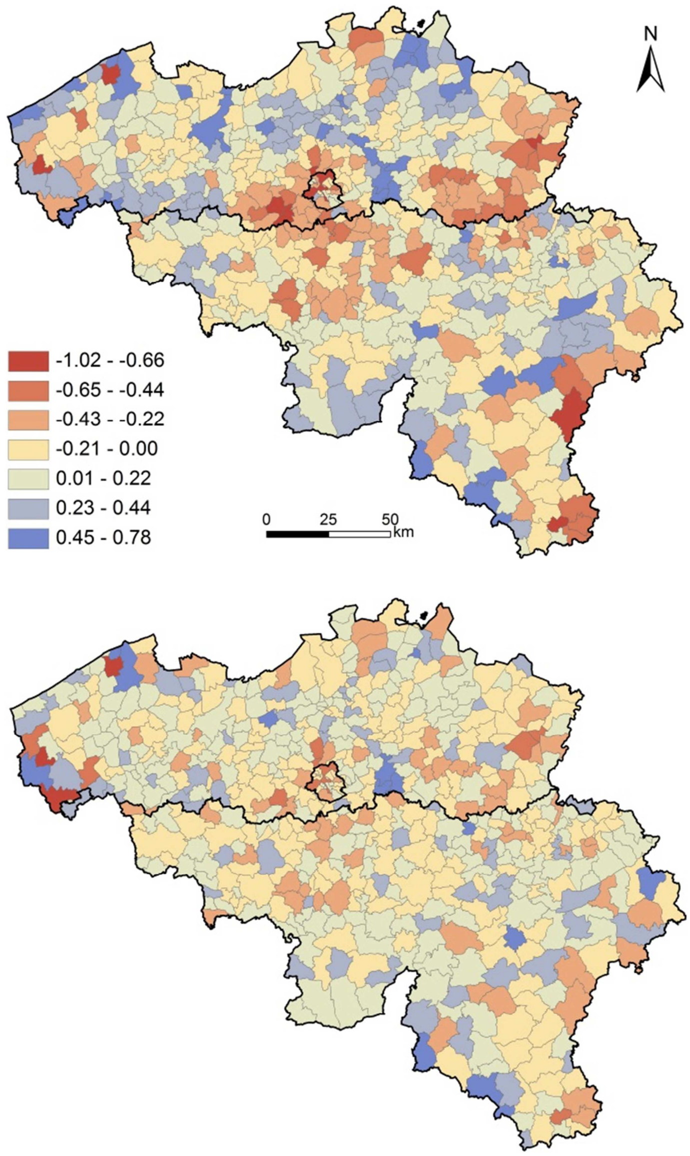


Figure 3.3: OLS (up) and ML residuals (down)

Table 3.3 shows that the spatial autoregressive coefficient ρ (or lag coefficient) is highly significant. This suggests that spillover influences exist between one municipality i and its neighbourhood: the likelihood of cycling in i is (positively) linked to bicycle use in the neighbouring municipalities j. The significance and magnitude of all the regression coefficients are lower for the ML estimation than for OLS, which can be explained by the introduction of the spatial autoregressive coefficient ρ. This suggests that part of the explanatory power of variables in municipality i may really be due to the influence of the neighbouring municipalities j (which is picked up by ρ). Among the significant coefficients, the average change in relative value is high (47%) and illustrates the substantial bias of the OLS model coefficients when spatial dependence is ignored.

The signs of the regression coefficients are the same for the OLS and ML estimations. As expected, most of the (significant) explanatory variables introduced in the models have a deterrent effect on the proportion of commuters cycling. Municipalities with high proportions of working people over 45, working households with one or more young children, or inhabitants in poor health have lower levels of commuter cycling. Municipalities characterised by large numbers of highly-educated people also have lower levels of commuter cycling, which confirms the results of previous studies in Belgium (Toint et al., 2001; Hubert and Toint, 2002). On the other hand, high levels of cycling are observed in municipalities with high proportions of working men. Among the environmental and policy-related variables, the presence of high accident risks, heavy traffic volumes and steep slopes along the road network are associated with a low propensity to cycle to work. The size of the town also matters, and this is probably associated with the provision of good facilities for cycling. The proportion of commuters cycling is the highest in the towns (well-equipped municipalities), and lowest in small municipalities. Interestingly, such results overall confirm those obtained in Chapter 2 (except for the largest towns, where the proportion of commuting by bicycle is lower than in most of the other towns for distances lower than 10 km).

As regards traffic volumes, opposite signs are unexpectedly observed between correlation coefficients (Table 3.1; positive correlations) and parameter estimates (Table 3.3; negative estimates). These peculiar results are explained by the fact that the relationship between independent variables is hiding/removing their true relationships with the dependent variable (Cohen et al., 2003). In such a case, the inclusion of an independent variable – called the 'suppressor variable' – in a multivariate regression model may enhance the results, in the sense it may remove the unwanted variance and/or reveal the true direct relationship between another independent variable and the dependent one (e.g. in leading to the expected sign in the regression, which suggests that the presence of (slight) multicollinearity may not be entirely undesirable) (Cohen et al., 2003; Friedman and Wall, 2005). In this chapter, the multivariate framework of the regression hence allows obtaining the true direct relationship between traffic volumes and cycle commuting (and then the right signs). Such an effect is here categorised as being a ‘negative net suppression’ one. Note that further information about the different types of ‘suppression effects’ is provided by Horst (1941), Darlington (1968), McNemar (1969), Conger (1974), Tzelgov and Henik (1981, 1991), Cohen et al. (2003), and Friedman and Wall (2005).

3.5.5 Accounting for spatial heterogeneity

3.5.5.1 Diagnostics: Chow tests and exploratory spatial data analyses

Structural instability is detected by the Chow test and its spatial extension (Table 3.2). Both tests (non-spatial and spatial) are highly significant and hence clearly reject the null hypothesis of parameter stability. This suggests that the spatial lag results do not completely account for spatial heterogeneity. Exploratory spatial data analyses confirm these findings. The global spatial autocorrelation for the dependent variable is first tested using the Moran’s I statistic. This latter is positive ($I = 0.90$) and highly significant ($p = 0.0001$), which indicates the presence of a positive and statistically significant degree of spatial autocorrelation in the distribution of y. Hence, this means that municipalities with low/high proportions of commuting by bicycle are generally located in the vicinity of each other.

Secondly, the Moran scatterplot and the LISA cluster map for the dependent variable help us to identify spatial regimes (Figure 3.4). The Moran scatterplot exhibits a positive spatial autocorrelation and suggests that the number of spatial outliers (or atypical municipalities) is low. Indeed, 95.1% of the municipalities fall either into quadrant I (HH; 42.6%) or quadrant III (LL; 52.5%) and hence show association of similar values. On the contrary, quadrants II (LH) and IV (HL) only account for 2.7% and 2.2% of the municipalities, respectively. The results of the Moran scatterplot hence suggest the presence of spatial heterogeneity in the form of two distinct spatial regimes, in quadrants I and III. The LISA cluster map illustrates the spatial pattern of these regimes and reveals a clear-cut north/south division of the municipalities: most of the northern municipalities (Flanders) fall into quadrant I, while a large proportion of the southern municipalities (Wallonia and Brussels) fall into quadrant III. Municipalities falling into quadrants II and IV are marginal (less than 0.5% of the significant LISA) and are consequently not considered as spatial regimes.

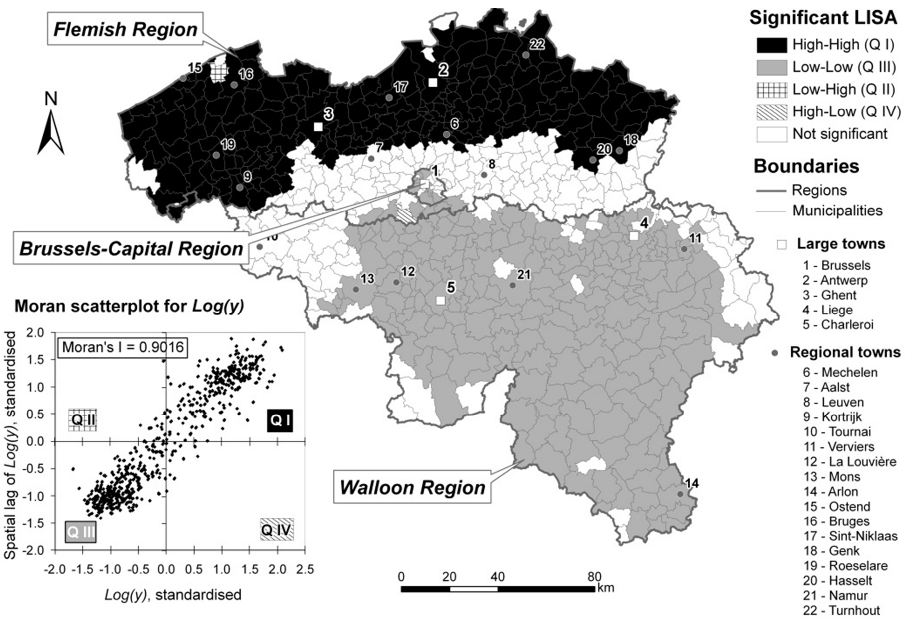


Figure 3.4: Moran scatterplot and LISA cluster map for the spatial clustering of commuting by bicycle. Note that significant LISA here refers to a 5% pseudo-significance level

3.5.5.2 Spatial regime regression with a spatially lagged variable

The northern and southern spatial regimes were incorporated into the regression to adjust for spatial heterogeneity. The diagnostics in Table 3.4 show the existence of spatial autocorrelation and heteroskedasticity in the models. This was corrected by applying an ML estimation (lag) and a White's correction (Table 3.5). The spatial regime specification (with spatial lag) gives a considerably better fit than the spatial lag model; the log-likelihood indeed increases from 33.7 (ML) to 93.9 (ML with regimes).

Several explanatory variables are significant for the north (Flanders) but not for the south (Wallonia and Brussels), and vice versa (Table 3.5). The signs of the significant coefficients are the same as in the spatial lag specification, but the magnitude differs greatly in some cases. For Flanders, the average change in the (relative) values ranges from 6.4% for dissatisfaction with cycle facilities to 426.5% for the accident risk. For Wallonia and Brussels, this change is less pronounced, ranging from 2.7% for the risk of an accident, to 58.7% for town size. These findings not only illustrate how biased the estimates are when the structural instability is ignored, but they also show the substantial difference in the size of these estimates between the Belgian regions (in this respect, the parameter estimate related to the risk of an accident is probably the most representative example).

Table 3.4: Regression diagnostics for the OLS and ML estimations, including the spatial regimes

	OLS with spatial regimes	ML with spatial regimes
Diagnostics for heteroskedasticity		
Breusch-Pagan test [1]	112.44***	94.46***
Koenker-Bassett test [1]	93.64***	79.74***
Diagnostics for spatial dependence		
Moran's *I* of residuals[2]	0.29***	-0.20
Lagrange multiplier (lag)	200.25***	n.a.
Robust LM (lag)	76.96***	n.a.
Lagrange multiplier (error)	130.33***	n.a.
Robust LM (error)	7.04***	n.a.
Diagnostic for spatial dependence and heteroskedasticity		
Joint LM test	242.77***	n.a.
Diagnostics for residual autocorrelation		
LM test	n.a.	0.16

*Significant at the 90% level; **Significant at the 95% level; ***Significant at the 99% level
n.a.: no test available
[1] Spatial version of the test
[2] Inference computation based on 9999 permutations (for ML estimation only)

3.5.5.3 Regional variation and the relative importance of the variables

Table 3.5 suggests that variables such as median income and the proportion of working men are not significantly related to the rate of cycle commuting in Wallonia and Brussels; on the other hand, they are positively related to the rate in Flanders. The positive relationship between median income and bicycle use can probably be explained by the fact that lower median income may act as a proxy for crime and vandalism (Parkin et al., 2008). This suggestion is supported by a significant correlation of –0.20 between the median income and the number of bicycle thefts in a municipality (and a correlation of –0.27 between median income and the risk of bicycle theft). The relationship between cycle commuting and the air pollution (the annual mean concentration of PM10) was also only significant in Flanders. Surprisingly, Figure 3.5 shows that, on its own, an increase in the PM10 concentration actually increases with the rate of cycle commuting in a Flemish municipality: for instance, an increase from 25 to 30 $\mu g/m^3$ is associated with a 8.1% increase of the bicycle share. This unintuitive result is rather difficult to explain, although it is here assumed that congested urban environments (which are generally areas with high concentrations of PM10) may play a role in explaining such a relationship. Indeed, increasing road congestion in urban areas probably encourages using other transport modes than car (e.g. the bicycle, which is faster than car during peak hours).

Three variables (education 3, bad health, traffic volume 2) which are not significantly related to bicycle use in Flanders do appear to have an impact in Wallonia and Brussels (the southern part of Belgium). The results suggest that a one percentage point decrease in the proportion of inhabitants reporting bad health will increase bicycle use by 0.07%. Relatively good physical and mental health is indeed required to use a bicycle. Moreover, a decrease from 25 to 15% in the proportion of highly qualified people in a municipality is linked to a 17.8% increase in commuter cycling[18]. Commuters with better qualifications generally get higher wages and fringe benefits such as a company car; this probably explains why they are more likely to have a car at their disposal, and so choose to live far from their workplace, beyond ‘cycling distance’. Finally, a reduction in the volume of motorised traffic is expected to encourage cycling: a decrease from 200,000 to 100,000 vehicle-km per kilometre of local road per year is predicted to increase bicycle use by 5.23% (Figure 3.6). Concretely, such a reduction corresponds to a 1,700-km decrease in the mileage of motorised vehicles per municipality per year, or a 4.6-km decrease in the daily mileage[19]. Ideally, a substantial reduction in traffic, from 1,000,000 to 10,000 vehicle-km (achieved, for example, through the implementation of an urban toll) could increase bicycle use by as much as 32%.

[18] Note that this result holds for Wallonia, but is not expected to be valid for Brussels since most commuter cyclists (66%) are here highly qualified. The increase in commuter cycling associated with a decrease in the proportion of highly qualified people is to be explained by the greatest weight of the Walloon municipalities in the spatial regime (262 municipalities, compared with the 19 Brussels municipalities).

[19] Assuming a 169-km local road network and 10,000 motorised vehicles using this network each year. These figures are based on the averages for Belgian municipalities.

Table 3.5: Regression coefficients for the spatial regime specification (ML estimation)

	ML, spatial regimes & heterosk. correction	
	North	*South*
Intercept	2.3084*	4.3095***
	[0.0000]	[0.0000]
Lag coefficient (ρ)	0.5362***	0.5362***
	[0,5097]	[0,5097]
Demographic variables		
% working men	0.0296**	0.0008
	[1.0246]	[0.0288]
% age 2 (45-54)	-0.0417**	-0.0205***
	[-0.5854]	[-0.3007]
% age 3 (> 54)†	-0.1074	-0.0680
	[-0.1317]	[-0.0867]
% young children (≤ 5 years)	-0.0365***	-0.0247***
	[-0.4372]	[-0.3306]
Socio-economic variables		
% education 3 (university degree)†	-0.0968	-0.3132***
	[-0.2104]	[-0.6862]
Income	0.0311*	-0.0027
	[0.3824]	[-0.0307]
% bad health	-0.0098	-0.0146**
	[-0.1274]	[-0.2481]
Environmental and policy-related variables		
Commuting distance	-0.0165***	-0.0047*
	[-0.2061]	[-0.0765]
Town size (urban rank)	-0.1146***	-0.0361***
	[-0.4539]	[-0.1483]
Slope†	-0.1931**	-0.1972***
	[-0.1145]	[-0.1966]
% dissatisfaction with cycle facilities	-0.0052***	-0.0045***
	[-0.1666]	[-0.2227]
Accident risk†	-0.7632***	-0.1489***
	[-0.1047]	[-0.0493]
Air pollution	0.0138***	-0.0054
	[0.2551]	[-0.0956]
Traffic volume 2 (municipal/local)†	-0.2357	-0.4521**
	[-0.0306]	[-0.0700]
N	589 (N_{North} = 308; N_{South} = 281)	
Log likelihood	93.923	
Akaike information criterion (AIC)	-123.846	
Schwarz information criterion (SIC)	16.264	

*Significant at the 90% level; **Significant at the 95% level; ***Significant at the 99% level
Standardised regression coefficients are given in brackets
†: logarithmically transformed variables

Most of the other explanatory variables have significant relationships with bicycle use in both regions. At first glance, variables such as the proportion of households with young children, or the proportion of commuters aged 45 to 54, seem to be strong deterrent factors for cycling. Their impact is also more pronounced in Flanders, than in Wallonia and Brussels. For instance, when a municipality increases its percentage of working households having one or more young children from 15 to 25%, the proportion of commuting by bicycle is reduced by 33.5% in Flanders, while it is reduced by 28.4% in Wallonia and Brussels. Similarly, an increase from 20 to 30% in the percentage of working people being between 45 and 54 years old results in reductions of 37.8% (Flanders) and 24.7% (Wallonia and Brussels) in the commuter cycling. Combined with the results of a principal component analysis based on an orthogonal varimax rotation not reported here), these findings also suggest that: (1) being young (i.e. less than 25 years of age) and having poor qualifications increases the propensity to cycle to work; (2) having more than one young child in the household increases the probability of owning a car, and consequently decreases the likelihood of cycling.

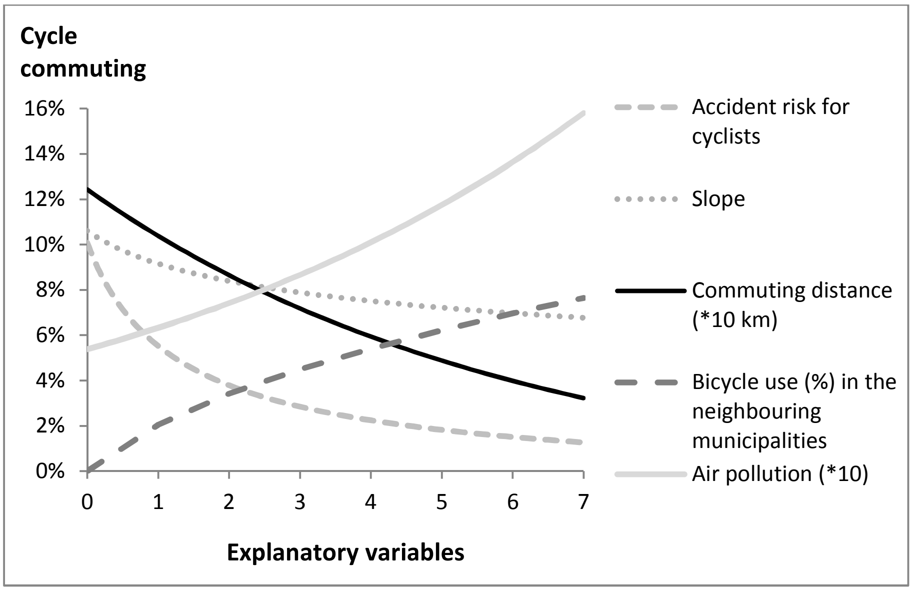


Figure 3.5: Variation in bicycle use in Flanders as explanatory variables change. Note: this figure is constructed by varying one explanatory variable, while holding all the others constant at their means. For ease of illustration, all the explanatory variables are all presented on the same *x*-axis. The sensitivity of the results was also tested for other values than the mean, i.e. the median, the lower quartile, and the upper quartile; except for the spatially lagged variable, such a sensitivity analysis suggests that our results are quite stable, whatever the chosen value.

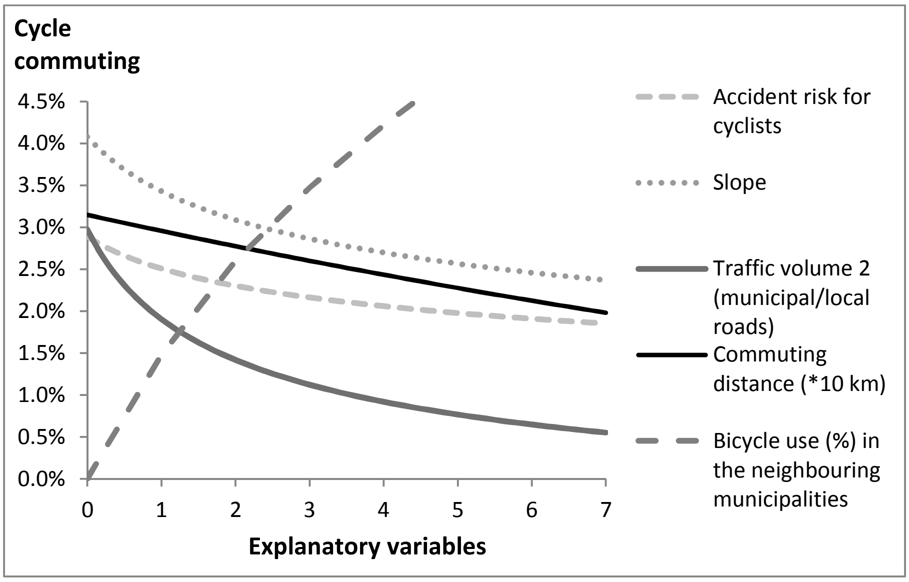


Figure 3.6: Variation in bicycle use in Wallonia and Brussels as explanatory variables change. Note: see note in Figure 3.5.

Table 3.5 also shows the impact that town size and dissatisfaction with cycle facilities have on cycle commuting. In both regions, a 10% increase in the proportion of households which express dissatisfaction with the facilities for cycling would reduce bicycle use by about 5.5%. Living in a poorly-equipped municipality (in terms of facilities) is associated with a lower likelihood of using a bicycle, whereas larger towns have a higher proportion of cycle commuting. In Flanders, the largest towns (H_1 to H_3) score well with cycle commuting rates 31.3 to 58.5% higher than in the most rural municipalities (H_6 to H_8). In Wallonia and Brussels, this difference is less pronounced, and only ranges from 3.4 to 9.7%.

Last but not least, Figures 3.5 and 3.6 suggest that a high risk of accidents, long commuting distances and hilly terrain decrease the propensity to cycle. Although it is basically similar in the two regions, the impact of the topography is slightly greater in the southern part of the country. An increase of the mean slope of the road network from 1 to 2° (which might occur, for example, when commuters are forced to take an alternative route, due to roadworks or deviations) could reduce the number of commuting cyclists by more than 8.4% in Flanders and 9.9% in Wallonia and Brussels. Conversely, reducing the slopes would significantly increase bicycle use, especially in municipalities where the mean slope is 1 to 2°. Above this 'limit', the impact of a change in the mean slope is lower. The risk of accidents also strongly discourages bicycle use. In Flanders, an increase in this risk from 0.0 to 0.5 (which corresponds to two more victims per year and per 100,000 bicycle-minutes) is linked to a 29.3% reduction in the number of commuter cyclists. In Wallonia and Brussels, the accident risk is also negatively linked to bicycle use, but to a lesser extent: the same increase in risk (from 0.0 to 0.5) is linked to a fall in bicycle use of only 7.9%. Longer commuting distances have a deterrent effect and do not stimulate cycling. In Flanders, an increase in the average commuting distance from 5 to 15 km would produce a 16.6% decrease in commuter cycling. The same increase in commuting distance in Wallonia and Brussels would reduce bicycle use by 6.1%. Given that the 'commuting distance' variable partly synthesises proximity-related information (through variables such as population and job densities, and urbanised areas), it implies that compact environments and tight town networks are associated with low commuting distances and hence stimulate cycling.

3.5.5.4 Spatially lagged variable

Table 3.5 shows that ρ is still highly significant and positive for the spatial regime model. This indicates the presence of a strong diffusion process between neighbouring municipalities: the (neighbouring) municipalities j exert a positive spillover effect on the propensity to cycle in municipality i, which in turn could generate a feedback effect on bicycle use in j. In the long-term, such a continuous diffusion process could initiate a 'mass effect', in the form of a virtuous circle which maintains the propensity to use a bicycle for commuting in the region. The municipalities in Wallonia and Brussels seem to be prone to a large reduction in bicycle use if cycling becomes less popular in neighbouring municipalities (and conversely), but Flemish municipalities are more resistant (relatively to Wallonia and Brussels) to the possibility of a fall in bicycle use in surrounding municipalities.

As suggested by Figures 3.5 and 3.6, the mass effect is more pronounced when municipality i and the neighbouring municipalities j all have low bicycle use (beyond 5%, it still increases, but to a lesser extent). This is even more true in Wallonia and Brussels: holding all other variables constant at their mean for this regime, an increase of commuter cycling from 1 to 4% in the neighbouring municipalities j will increase bicycle use from 1.5 to 4.2% in municipality i (= 184.8% increase), and next will initiate a feedback effect maintaining a continuous increase in municipalities i and j. In Flanders, the same increase is observed, but to a lower magnitude (+164.2%). Interestingly, such neighbourhood processes are also encountered in the reality. Figure 3.7 is quite evocative in this sense since it shows that variations in commuter cycling (1991-

2001) are spatially clustered in Belgium. For instance, Brussels and its periphery experienced an increase of bicycle use between 1991 and 2001, whereas reductions in commuter cycling were observed for clusters of 'rural' municipalities. The combination of a range of socio-economic factors (e.g. rising income, higher car availability, and larger commuting distances) probably explains such trends in bicycle use.

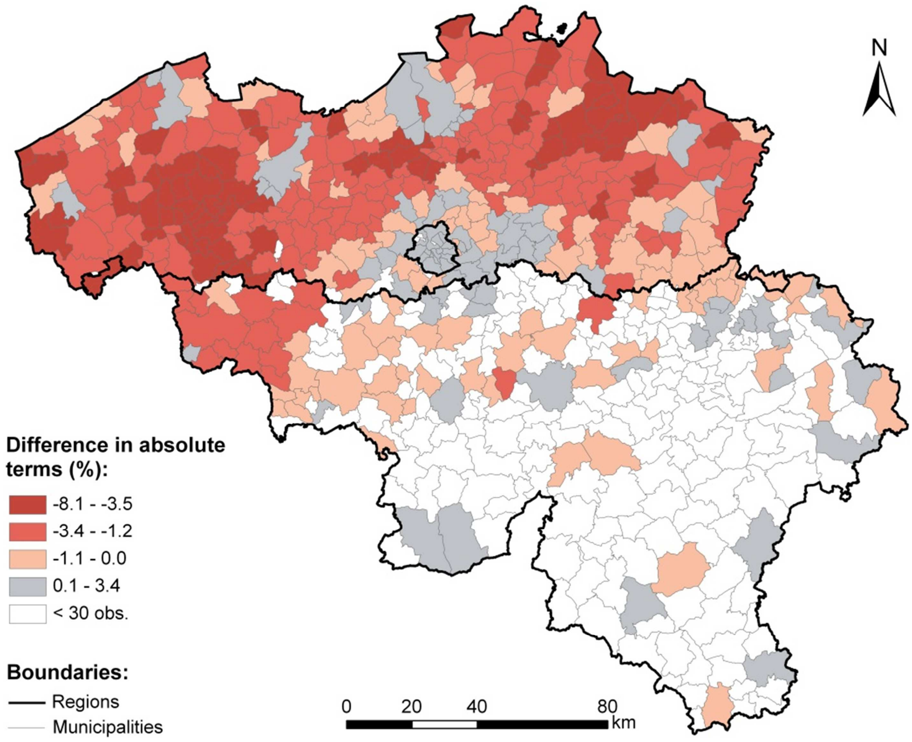


Figure 3.7: Absolute difference in commuter cycling between 1991 and 2001 (source: Verhetsel et al., 2007)

3.5.5.5 Analysis of the residuals

The residuals of the final specification (Table 3.5) are mapped in Figure 3.8. This provides a useful tool for planners and policy makers since it pinpoints both the municipalities that 'over-perform' in terms of bicycle use and those where there is still potential to develop the use of the bicycle for commuting trips further. This potential exists in the municipalities characterised by negative residuals. Given the current environment, such municipalities could perform better in terms of bicycle use but, for something (e.g. an inadequate or unambitious cycling policy, high-quality public transport) holds it back.

Examples of municipalities exhibiting negative residuals are Antwerp, Brussels or Genk. Gent and Kortrijk are also highlighted in Figure 3.8. It could be more surprising in view of their voluntary cycling policies and the high proportions of commuting by bicycle, although it may suggest that there is still some potential to get more cyclists here.

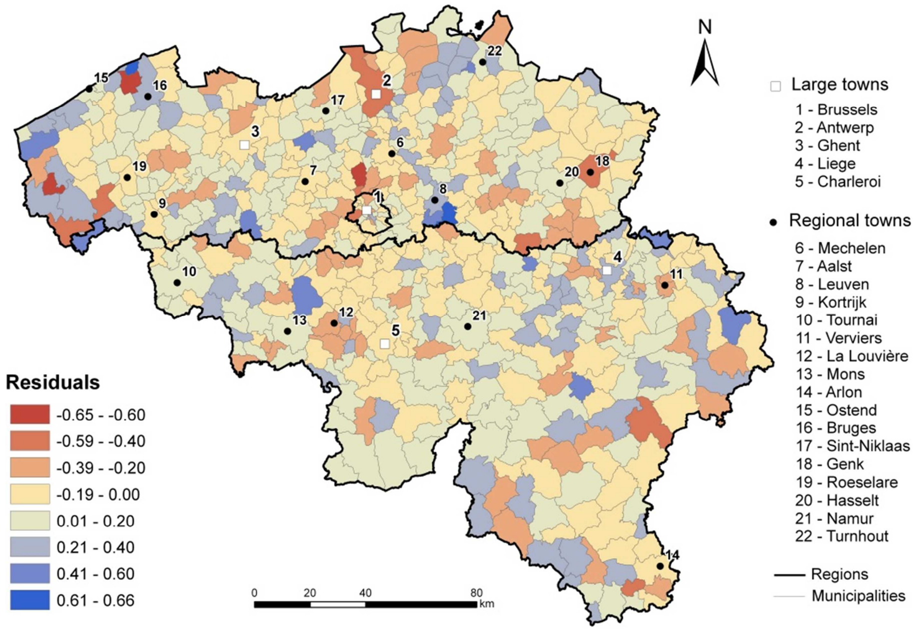


Figure 3.8: The residuals of the spatial regime specification (see Table 3.5)

At the other end, municipalities characterised by positive values of the residuals excel in terms of bicycle use (given their environment). The examples of Louvain and Bruges are important in this respect, since they have more pro-cycling policies (in terms of engineering, traffic education and enforcement) than other Flemish municipalities. Several municipalities in Wallonia also perform better than expected, despite their low absolute rates of cycle commuting. Given their environment (steep slopes, rural setting, etc.), they 'over-perform', for example by adopting mobility strategies that encourage bicycle use (SPW, 2008). Examples of such 'over-performing' municipalities are Ottignies-Louvain-la-Neuve, Perwez, Hotton, Yvoir, Marche-en-Famenne, Tournai and Mouscron.

3.5.6 Controlling for short commuting distances and spatial interactions

3.5.6.1 Effect of short commuting trips (≤ 10 km)

When the regression is carried out on the proportion of cyclists among commuters who travel less than 10 km (in municipality *i*), the results are basically similar to those shown in Table 3.5 (see Appendix B.2 for further information about the parameter estimates). The main difference lies in the variable referring to commuting distances: for commuting trips of up to 10 km, increasing distance is linked to more cycling, whereas in the general regression (Table 3.5) increasing distance is linked to less cycling. As revealed in Chapter 2, cycling is a very convenient mode of transport for distances between 2 and 5 km, but for shorter distances (0–2 km) walking is the preferred mode of transport. This suggests that, up to 10 km, commuting distance does not act as a strong deterrent to cycling. Given that approximately 39% of commuters (and even more in urban areas) live less than 10 km from their work, there is considerable potential for a shift to cycling.

3.5.6.2 Effect of spatial interactions (direct, indirect and total impact estimates)

Some additional statistical analyses were carried out in Appendix B.3, with the aim to check the magnitude of the impact associated with the presence of complex spatial interactions in the model (here referred as 'direct' and 'indirect' effects on cycling levels). Such interactions may indeed affect the validity and the interpretation of the results. They are here referred as 'direct' and 'indirect' effects on cycling levels. According to LeSage and Fisher (2008) and Lesage and Pace (2009), '**direct effects**' on cycling levels in municipality *i* may arise from a change in a single explanatory variable in *i*; these include: (1) the effect of a change **through** *i* (i.e. without considering the neighbourhood), and (2) feedback influences resulting from impacts (caused by the changes in explanatory variables in *i*) passing through the neighbouring municipalities *j*, and coming back to the municipality *i* itself (feedback loop). '**Indirect effects**' (or spillover effects) on cycling levels in *i* may also arise from changes in all the neighbouring municipalities *j* of an explanatory variable (LeSage and Fisher, 2008; Fisher et al., 2009; Kirby and LeSage, 2009; LeSage and Pace, 2009).

Analyses conducted in Appendix B.3 hence aim at evaluating the validity of our results reported in Table 3.5, controlling for the presence of direct and indirect effects in the model. They fortunately confirm the validity of our results. Given that the theoretical framework related to direct/indirect effects goes beyond the scope of the methodology described here, readers are advised to read Appendix B.3 if they wish to obtain further details about these statistical analyses.

3.6 Conclusion

The objective of this chapter was to explain the variation of the proportion of commuters who travel by bicycle at the scale of the Belgian municipalities. It then aimed at providing statistically-based recommendations in order to support planners and policy makers to initiate a shift from car to bicycle among commuters (see part IV, Chapter 6). In line with the literature on transport mode choice, our results suggest that demographic and socio-economic variables significantly influence the proportion of commuting by bicycle. Income, age and gender have a significant impact on the rate of cycle commuting in Flanders: low median income, low proportions of working women, and a young (under 45) workforce are all associated with high rates of cycling to work. Having one or more young children (less than 5 years old) in the household decreases the likelihood of cycling to work in both regions. The presence of many highly-qualified people also matters, particularly in the southern periphery of Brussels. Highly qualified commuters living in Wallonia and having high incomes, can afford a car, and use it to travel large distances. They are hence less likely to use a bicycle for their commuting trips (Jensen, 1999).

Furthermore, this chapter confirms the significant impact of several environmental and policy-related variables on bicycle use. Overall, municipalities that are well-equipped (i.e. large and regional towns) and characterised by short commuting distances have high rates of commuter cycling. Large urban areas indeed provide high-quality public transport and benefit from the proximity of different activities and the good connectivity between them, so that commuting distances are shorter and more bikeable. Flat terrain, high-quality cycle facilities and a low risk of accidents can also encourage commuter cycling in both regions. However, heavy traffic (on municipal roads) does not have any significant impact in Flanders, whereas it strongly discourages cycling in Wallonia and Brussels. In Flanders, the high visibility of cyclists in the traffic (because there are such a lot of them) and the presence of appropriate cycle infrastructures probably give commuter cyclists a feeling of personal security and, hence, offset the deterrent effect of traffic volume. Policies in Flanders do indeed provide high-quality infrastructure (e.g. continuous and separated cycle lanes) and facilities (e.g. changing facilities at work) with the intention of improving the safety and convenience of cycling. Flanders also stimulates bicycle use through regulations restricting motorised traffic in urban centres (e.g. through the introduction of traffic calming areas), so that the risk and annoyance of heavy traffic is greatly reduced. Finally, motorists show more respect for cyclists because they often cycle themselves and/or are used to sharing the road with large numbers of cyclists. The opposite situation is observed in Wallonia and the Brussels region: here, the terrain is hillier and discourages cycling. Also, motorists are seldom mindful of commuter cyclists and still consider them less important than car drivers (especially in Wallonia). Due to a lack of cycle infrastructure in the Walloon municipalities, the risk of being seriously injured or killed is high (especially in rural areas), and confirms residents' fears of cycling. This is not, however, the case in Brussels, where the risks of severe/fatal accident for cyclists are low. Chapter 2 indeed suggested that the urban environment, with its large number of obstacles, forces drivers to reduce their speed.

From the methodological point of view, the modelling techniques applied here highlight the importance of accounting for multicollinearity, spatial dependence and spatial heterogeneity (i.e. structural instability and heteroskedasticity). Spatial autoregressive models appear to be very powerful in eliminating spatial autocorrelation, while the presence of spatial heterogeneity in the data is corrected using White's correction and a spatial regime regression. More interestingly, the presence of spatial dependence in the model suggests that bicycle use in a municipality is influenced (positively or negatively) by the neighbouring municipalities, i.e. a municipality surrounded by others with high levels of cycling is more likely to show high rates of commuter cycling (and vice versa). This indicates that social support for cycling could stem from the neighbourhood. This confirms results obtained at a more disaggregated scale (de Geus, 2007; de Geus et al., 2008b). Besides spatial dependence, the need to adopt a spatial regime specification indicates that different effects exist in the northern (Flanders) and southern (Wallonia and Brussels) parts of Belgium. Pro-cycling strategies should hence be approached from different strategies, without however neglecting inter-regional exchanges since these are crucial to learn from each other (in terms of experience) and to develop a constructive approach with respect to non-motorized modes of transport.

Part III: Spatial analysis of accident risks for cyclists (Brussels-Capital Region)

Chapter 4

Reported *versus* unreported cycling accidents A spatial network analysis for Brussels[1]

Outline

In Belgium as in most countries, a large share of cycling accidents (> 85%) is not registered by the police and then does not appear in official statistics of road accidents. Cyclists involved in these 'unreported' accidents generally incur slight injuries and/or material damages, and are often the single road users involved in the accident. Hence, they often do not feel the need to call the police. This chapter then aims at providing further knowledge in the hidden part of cycling accidents and focuses on the Brussels-Capital Region (Belgium), which is here subdivided into three 'subareas' (i.e. Pentagon, First and Second Crowns). The main objective is to explore and compare the spatial patterns of cycling accidents registered by the police with those unreported (by the police) but collected through an open-based online registration survey (SHAPES survey). It also aims at analysing whether or not unreported and reported cycling accidents have similar locational tendencies with respect to specific road infrastructures. Comparative statistics, point pattern exploration techniques and (cross) K-function methods are here applied into a Geographic Information System (GIS) and – when possible – extended to the road network using a GIS-based extension

[1] This chapter will be submitted in 2011 for publication. It is adapted from: Vandenbulcke, G., de Geus, B., Thomas, I., Aertsens, J., Meeusen, R., Int Panis, L. Reported *versus* unreported cycling accidents: a spatial network analysis for Brussels.

(SANET v.4). Our findings reveal that, for a given subarea, reported and unreported cycling accidents have similar spatial patterns and overall exhibit similar locational tendencies with respect to specific infrastructures and facilities. Methodologically, it is also demonstrated that the results of (cross) K-function methods depend on the chosen spatial subarea and, hence, should be interpreted with great caution. Last but not least, we show that cycling accidents are more prone to be unreported by the police in areas where a lower differential between the speed of slow and fast modes is imposed (e.g. traffic-calming areas). Such areas indeed lead to accidents with a lower degree of injury severity, which then reduces the need to call the police and – in turn – may conduct to higher rates of underreporting.

4.1 Introduction

In most countries with car-oriented policies, the fears and safety concerns about on-road cycling are high. Making cycling trips safer is hence an essential step for encouraging more and more people to cycle and, then, for contributing to health, environmental and mobility policies. Besides strategies focussed on enforcement (e.g. police controls), traffic education (e.g. through awareness campaigns) and encouragement, the detection and analysis of locations where cycling accidents spatially concentrate along the network (i.e. 'black zones') play a prominent role in pinpointing *where* investments in road infrastructure modifications should have priority to enhance bicyclists' safety. Also, it turns out to be helpful in suggesting causal relationships between cycling accidents and specific factors (e.g. infrastructure-related factors) (Flahaut et al., 2003; Steenberghen et al., 2004). Nevertheless, in many countries, it is well-known that cycling accidents with slight injuries (and/or with material damages) are strongly underreported compared to other degrees of severity. In Belgium, several authors estimate that about 15% of the cycling accidents are reported in official statistics (see e.g. Hubert and Toint, 2002; Lammar and Hens, 2004; Doom and Derweduwen, 2005; De Mol and Lammar, 2006; Lammar and Hens, 2006). As a corollary of such a poor data registration, the identification of black zones for cyclists (and the underlying factors associated with cycling accidents) might be biased or inaccurate, especially if 'unreported' accidents (i.e. accidents that are not registered by official/governmental agencies) exhibit different spatial and/or temporal patterns compared with the reported ones.

To our knowledge, no research has been conducted to get insight in the spatial patterns of unreported cycling accidents, and nothing is known in the literature about the (possible) spatial differences between these latter and the reported accidents (i.e. those compiled by official/governmental agencies). Given that unreported accidents are by far the most numerous (among all accidents) and that they mainly consist of single-vehicle accidents with slight injuries and/or material damages, it could be worth to question/analyse if they occur at different locations and if such locations are characterized by different infrastructure factors, compared with the reported ones.

Hence, this chapter aims at: (1) exploring and comparing the spatial patterns of cycling accidents registered by police with those unreported by police but collected through an online registration survey (SHAPES survey; see Aertsens et al., 2010; Int Panis et al., 2011; de Geus et al., accepted); and (2) inspecting if reported and unreported cycling accidents both cluster around similar spatial factors/variables (or, in other words, if similar neighbourhoods are at the root of both reported and unreported accidents). Such analyses are conducted within the Brussels-Capital Region (BCR). They not only make use of comparative statistics and measures of central tendency and dispersion, but also take advantage of the use of point pattern methods extended to networks, such as the network *K*-function and cross *K*-function methods (see e.g. Okabe and Yamada, 2001; Okabe et al., 2006a, 2006b).

Interestingly, differences in the spatial patterns and (accident-related) factors would be indicative of the fact that unreported and reported cycling accidents locate at different places along the network and hence that explanatory variables are probably neglected when focussing on reported cycling accidents only. It then suggests that a more complete registration of cycling accidents would provide additional and/or more accurate information about the significance of spatial factors associated with the occurrence of cycling accidents. At the opposite, no or little difference between reported and unreported accidents would suggest that improving the accident registration (e.g. through surveys) would not necessarily provide additional information on unobserved spatial factors, although it is here thought that a more accurate (spatial) representation of black zones for cyclists would be helpful in pinpointing the locations where local safety treatments are needed to improve the bicyclists' safety.

This chapter is structured as follows. Section 4.2 introduces the studied area (Brussels-Capital Region) and provides some figures in terms of bicycle use and accident risks for cyclists. Section 4.3 describes the data, after which the methodology is presented in Section 4.4. Results are reported and discussed in Section 4.5. Finally, Section 4.6 concludes this chapter by summarizing its main findings and limitations.

4.2 Spatial context: the case of the Brussels-Capital Region

4.2.1 Diagnosis: mobility and cycling levels in Brussels

Centrally located in Belgium, the BCR is a highly urbanised area, where more than 1,125,000 inhabitants concentrate over 162 km² (population density is hence about 7000 inh./km²). As capital of Belgium and Europe (EU-27), this region concentrates lots of facilities (e.g. administrations, head offices, transport, education, etc.) and is hence a major area of employment in Belgium, generating more than 650,000 jobs and 20% of the national GDP (Thisse and Thomas, 2010). Such a concentration of population and activities – combined with a high level of accessibility of/to most of the transport networks – results in a high attraction of the region, and then high traffic volumes. About 700,000 trips (all purposes) are registered every day in the region during the morning peak hours (6-10 a.m.). Among these, 64% are carried out by car, which is mostly explained by the high motorisation rate of the region (i.e. one vehicle for less than 2 inhabitants) and the continuous urban sprawl in the peripheral municipalities that generate every day more and more car trips converging on the capital (Dujardin et al., 2007; Brussels Mobility, 2010; De Witte and Macharis, 2010).

In order to mitigate car-related externalities, the Brussels' mobility plan (IRIS II) intends to reach a 20% decrease in car traffic by 2020 (in vehicles/km, using the year 1999 as basis). Encouraging a modal shift from car to active modes of transport − besides promoting public transport and rational car use − is one of the measures proposed by the plan in order to achieve such a target. Although the number of cyclists strongly increased in Brussels since 2000 (Figure 4.1), the share of cyclists in the traffic is still low compared to other European towns (about 4%). There are indeed important barriers that deter people from cycling in Brussels, such as the high motorised traffic levels with which the cyclist can conflict and the (perceived) risk of being involved in a cycling accident. Despite the fact that the risk of being killed or seriously injured is relatively low in Brussels, Figure 4.1 shows that the risk of cycling accidents in general remains quite stable since 2002 due to the simultaneous increase in the number of accidents (BRSI, 2009). This hence somewhat confirms the resident's fears about the risk of being involved in a road accident when cycling.

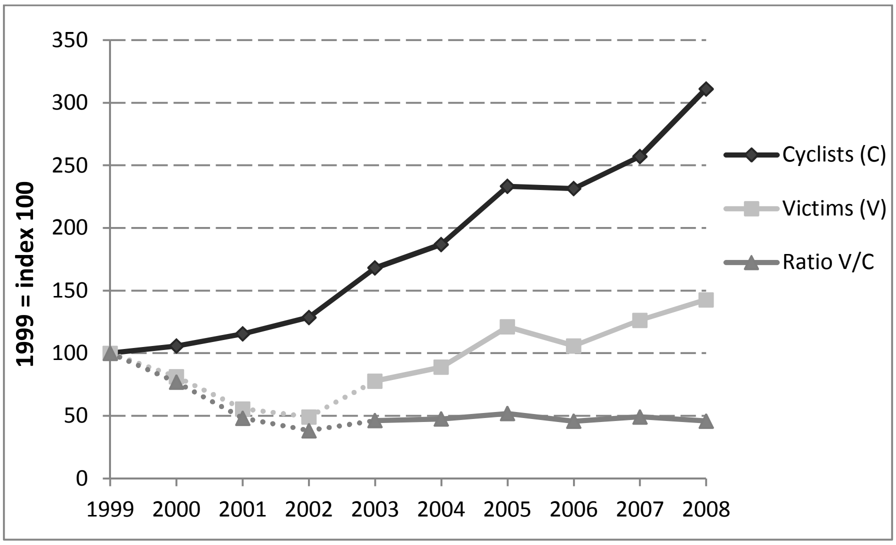


Figure 4.1: Evolution of the average number of cyclists (C), number of victims of cycling accidents (V), and ratio V/C. Dotted lines = period with strong under-registration of road accidents (1999-2003). Data sources: BRSI, 2009; DGSEI; Pro Velo, 2011.

4.2.2 Why Brussels?

The BCR is an interesting case study for several reasons : (1) most of the transport policies and planning decisions (e.g. provision of cycle facilities, traffic-calming measures, parking policies, etc.) are conducted at a regional scale in Belgium; (2) it is a highly urbanized area, characterised by a relatively high number of (reported and unreported) cycling accidents compared with rural areas, which hence increases the significance of the results and the probability to identify hot spots of cycling accidents; (3) due to the urban context, most of cycling accidents (95%) occurring in Brussels result in slight injuries, which hence provides a rather homogeneous accident data set in terms of severity; (4) the spatial variability of some spatial factors is quite large (e.g. the number of roundabouts increases while moving away from the city centre; most of the cycle facilities are located close to the European and Regional institutions; etc.); and (5) a wide range of data and information about cycling and the potential factors being at the root of cycling accidents are available for the BCR (through e.g. aerial photographies, digitized data, cycling maps, etc.).

However, there is also some inconvenience to focus on the BCR only. Among the main limitations, the fact that the studied area and the data are limited by administrative/regional boundaries (instead of socio-economic boundaries) is expected to cause edge effects of the first type. For instance, it might hamper the ability of the network cross-K function methods to detect a significant clustering (or dispersion) of cycling accidents around some definite 'peripheral' factors, i.e. factors that are more likely to be located in the periphery of the extended urban agglomeration (i.e. the area defined by socio-economic and demographic criterions) rather than in the Central Business District (CBD).

4.2.3 Spatial subareas

Within the framework of this chapter, the BCR is subdivided into three zones in order to examine the impact of increasing spatial subareas on the results (see Figure 4.2 in Section 4.5.2). Firstly, the 'Pentagon' (4.5 km^2) is the centremost part of the BCR and corresponds to the Brussels' historic city centre. It is pentagon-shaped and is delineated by an inner ring road (called the Brussels small ring) that is built on the site of the second set of defensive walls of the city (16th century). It includes districts with high densities of jobs (15,000 jobs/km^2) and population (9,000 inhab./km^2), and attracts everyday high cycling flows that mainly come from the First Crown (and, to a lesser extent, from the Second Crown). Secondly, the First Crown (39.1 km^2) designates the districts situated between the inner ring road (or Pentagon) and the greater Brussels ring. This latter consists of a set of major boulevards (and railways in the western part) that are intermediate between the Brussels small ring and the main ring (motorways). It also surrounds districts characterised by high densities of population (12,000 inhab./km^2) and jobs (4,000 jobs/km^2), and built before 1914. Lastly, the 'Second Crown' (118.9 km^2) corresponds to the area situated between the greater Brussels ring and the administrative boundaries of the BCR. It includes districts built during the 20th century and for which the population and job densities are generally lower (the densities raise to 4000 inhab./km^2 and 1000 jobs/km^2, respectively), compared with the Pentagon and the First Crown.

Our spatial point pattern analyses are carried out assuming that a definite spatial subarea includes any other embedded spatial subarea. For instance, it means that the First Crown here refers to all subareas it embeds (i.e. First Crown *+ Pentagon*). Similarly, the Second Crown corresponds to the whole BCR (i.e. Second Crown *+ First Crown + Pentagon*). Note that the use of increasing sizes of subareas here aims at monitoring the effect increasing numbers of observations have on the results; it is not the aim here to isolate the characteristics of the different subareas.

4.3 Data collection

Spatial analyses on networks are here performed into Geographic Information Systems (GIS), taking advantage of the availability of a free ArcGIS-based extension called 'SANET' (Spatial Analysis on a NETwork)[2]. Data collection is hence carried out into ArcGIS (through a digitizing process) and here consists of a three-step approach: (1) a 'bikeable' network is defined and constructed over the entire study region (Section 4.3.1), (2) this network is then used as reference material (in address matching techniques) for geocoding the reported and unreported cycling accidents (Section 4.3.2), and (3) some of the main (spatial) factors associated with the presence of cycling accidents are reviewed, digitized into a GIS and, then, transformed into point features when necessary (Section 4.3.3).

4.3.1 Construction of the 'bikeable' network

The road network for the BCR is provided by the Brussels Regional Informatics Center (BRIC), using the Brussels UrbIS database. Among the 2137 km of roads included in the Region, approximately 120 km are excluded because they are 'unbikeable', i.e. they are forbidden to cyclists or not designed to accommodate bicycle traffic. Orthophotos for the years 2004, 2007 and 2009 (BRIC, Google Earth) and cycling maps for the 2006-2008 period (Brussels Mobility) are used to identify and exclude such 'unbikeable' links. Overall, these latter are motorways and parts of the network without any cycle facility (e.g. slip and access roads, express roads, bridges, tunnels). The remaining 2017 km of road links hence correspond to the so-called 'bikeable network'. Modelling such a network into a GIS allows computing network distances (instead of Euclidean distances) between points located along the 'bikeable' network. This provides a good estimation of the spatial relationships existing between network-constrained points (see e.g. Yamada and Thill, 2004; Okabe et al., 2006a, 2006b; Shiode, 2008; Steenberghen et al., 2010).

Note that the three spatial subareas considered here include different network lengths, different databases (e.g. concerning the number of accidents occurring on the network), and – as a corollary – different computation times into SANET. Considering a 10m buffer for each of these spatial subareas (in order to mitigate the edge effects as much as possible), the total length of the street network

[2] This plug-in tool has been implemented in ArcGIS 9.3 by a group of Japanese researchers (Okabe et al., 2006a, 2006b, 2009), with the aim to operate network methods in GIS. SANET v.4 beta is here used.

amounts to 110 km, 791 km and 2030 km for the Pentagon, First Crown and Second Crown (respectively).

4.3.2 Accident geocoding

4.3.2.1 Reported cycling accidents (DGSEI data) and the under-registration issue

In Belgium, road casualties are registered by the police and compiled annually by the Directorate-General Statistics and Economic Information (DGSEI). In this chapter, a total of 644 bicycle accidents are censused over the period 2006-2008 and for the whole studied area (BCR)[3]. The severity of the accident is not considered here, but this should not be a major concern since few serious injuries (25) and no fatality were reported in Brussels during the period of study. At the opposite, 95% of the cyclists involved in a road accident suffered only slight injuries. This last figure is even expected to be higher since bicycle accidents with slight injuries (and/or with material damages) are strongly underreported compared to the other degrees of severity. Such underreporting is explained by the fact that bicycle accidents are often single-vehicle accidents, characterized by minor injuries and/or material damages. In such cases, the cyclist generally does not feel the need to call the police (and hence there is no official record) and cures oneself and/or repairs oneself the material damages (BRSI, 2008, 2009).

4.3.2.2 'Unreported' cycling accidents (SHAPES survey)

An open-based online registration survey was implemented within the framework of a Belgian research project (SHAPES) in order to get better insight into minor/slight cycling accidents (i.e. location, costs, underreporting, etc.) and the factors related to their occurrence in Belgium (Aertsens et al., 2010; de Geus et al., accepted). Within the scope of this chapter, such a survey was helpful to extract 'unreported' cycling accidents since it registered a large share of cycling accidents that were not reported by the police (and, hence, by DGSEI). In Brussels, a comparison between the survey data (SHAPES) and DGSEI data indeed highlighted that only 7% of recorded cycling accidents were officially reported by police in the period from March 10th 2008 until March 16th 2009.

[3] Note that 3 bicycling accidents were added in the total of bicycling accidents since they were initially supposed to have occurred in the Flemish Region (according to the description of the accident by the police).

Although open-based, recruitment of the participants that registered on the survey was based on the following inclusion criteria: (1) age between 18 and 65; (2) cycling to work at least twice a week during the preceding year; (3) having a paid job outside home; (4) living in Belgium. A cohort of 1187 participants was then obtained based on such criteria and after a one-year follow-up period (March 10^{th} 2008–March 16^{th} 2009). Every week, each of these participants had to fill out a travel diary in order to report information on bicycle usage (i.e. on trip purpose, frequency, time and distance) during the preceding week. If a cycling accident occurred during the weekly registration, then the participant was automatically asked to fill out a 'prospective questionnaire' in order to register detailed information about the accident (about e.g. the circumstances, the cause of the accident and injury, the registration by police, etc.). One week after having received the first travel diary, a retrospective questionnaire was also sent to the participants to register the eventual cycling accidents they incurred during the preceding year (i.e. from March 10^{th} 2007–March 9^{th} 2008). As a result, a two-year period was then covered by the survey as regards the registration of cycling accidents. In the case where cyclists provided incomplete or erroneous information about the accident location, they were contacted once again in October 2009 and asked to pinpoint in Google Map the exact location of their accident. As a result of this prospective and retrospective registration, a total of 55 bicycle accidents is registered over the period from March 10^{th} 2007 until March 16^{th} 2009 and for the entire BCR. Eliminating the cycling accidents registered by the police from this total, the number of (unreported) cycling accidents then amounts to 51 (which corresponds to 93% of the accidents registered by the SHAPES survey). Interestingly, for either the BCR or Belgium, accidents that are not registered by the police involve slighter injuries for the cyclist (i.e. mainly material damages, bruises and/or cramps) than these registered by the police during the survey (which led to body injuries with short- or long-term consequences) (Aertsens et al., 2010). For further information about the survey and SHAPES, see Aertsens et al. (2010), de Geus et al. (in prep.) and Int Panis et al. (2011).

4.3.2.3 Accident geocoding process

Reported (DGSEI) and unreported (SHAPES) cycling accidents are separately geocoded using address matching techniques in GIS. Basically, the geocoding process requires two types of information: (1) the ***accident data***, which contain detailed information on the location of the accident (i.e. the municipality code, the street name(s) and the house number in front of which the bicycle accident occurred), and (2) the ***reference data***, i.e. the 'bikeable' network and house numbers (BRIC), which are both available in spatial formats and contain address elements that are compatible with accident data. After some preliminary steps (i.e. specifying the geocoding options, formatting the data, etc.), accident data are then automatically matched with reference data and assigned x,y coordinates along the network (ESRI, 2010). In order to improve the precision of the geocoding process, the bicycle accidents occurring at junctions are matched with the bikeable network, whereas the bicycle accidents occurring along the streets (i.e. between the junctions) are matched with the house numbers, before being snapped to the closest point of the network. Finally, using orthophotos for the years 2004, 2007 and 2009 (BRIC, Google Earth) and network data (BRIC), we manually checked the validity of the results obtained through the automatic geocoding process and tried to geocode the bicycle accidents that were not located due to the presence of spelling errors or incomplete information in the data. As a result, 93% of the officially reported cycling accidents (= 600 properly geocoded / 644 DGSEI accidents) and 96% (= 49) of the 51 unreported accidents (survey-based) were successfully geocoded in a GIS.

4.3.3 Infrastructure factors

Infrastructure factors are collected in order to explore their (expected) relationships with the occurrence of reported and unreported cycling accidents. Although road accidents generally result from the interaction and combination between five categories of factors (driver behaviour, vehicles, infrastructures, traffic conditions and environment) (Miaou et al., 2003; Li et al., 2007; BRSI, 2008), we only focus on infrastructure factors since most of these have a spatial dimension (e.g. in the form of x,y coordinates) whereas it is not always the case forthe other factors. Infrastructure factors are digitized either as linear objects (e.g. cycle facilities) or as point objects (e.g. public transport stops). In the case where they are linear-shaped, the factors are summarized as *centroids* in order to make possible the use of the above described point pattern methods. Concerning the point objects, no conversion is required since they are already digitized in a 'usable' format for point pattern analyses.

Note that a review of the literature is carried out in the next subsections as regards the infrastructure factors and their impact on the number of cycling accidents and – more particularly – on the risk of cycling accidents and injuries. Appendix C.1 also lists and describes all infrastructure factors used in this chapter. All of these data are digitized using one of the following sources: orthophotos (BRIC, Google Earth), printed maps (Brussels Mobility, City of Brussels), accident data (DGSEI), on-line applications (BRIC), or GIS data coming from the Brussels UrbIS database (BRIC) and STIB/MIVB (as regards tram infrastructures). These data are collected for the period 2006-2008 and at the scale of the Brussels-Capital Region (considering the three spatial subareas separately). The list of infrastructure factors is quite exhaustive since we here aimed at monitoring if some of these factors might have had an unexpected spatial relationship with the occurrence of cycling accidents. However, some factors are deliberately ignored due to frequent infrastructure changes, or simply because detailed data are difficult to obtain.

4.3.3.1 Bridges and tunnels

Bridges and tunnels are expected to be 'black spots' for cyclists because sudden change may sometimes occur here in terms of infrastructures and road conditions (Khan et al., 2009). They are the main crossing points of specific hurdles (e.g. rivers, railways, motorways) and a wide range of transport users often concentrate and share the road at these locations. The space devoted to each mode is hence reduced, forcing the road users to adapt their driving behaviour to the road environment. In particular, bridges are elevated infrastructures that may decrease the long-distance visibility of road users, e.g. due to the curving/bending. Given that they are seldom surrounded by buildings (due to their elevated position), they are more likely to be exposed to 'extreme' weather conditions. For instance, in the case where they cross rivers or water zones, bridges are places more prone to ice development (and hence road accidents) when low temperatures, strong winds and water evaporation occur jointly (see e.g. Khan et al., 2009). Finally, tunnels and road sections located below elevated infrastructures (e.g. road bridge, railways) force the cyclists and the other road users to adapt their eyes to the lower luminance level, hence increasing the perception time and the risk of having an accident (Wang and Nihan, 2004). In this chapter, only bridges with safeguards on both sides are considered; those surrounded by buildings and protected from variations in weather conditions are not selected. Note that tunnels prohibited to cyclists are not selected here.

4.3.3.2 Traffic-calming areas

Speed-related accidents are expected to be more severe and greater in number on roads where speed limits are high (e.g. 70 km/h, or more). Indeed, high speed is not only related to accident risk, but also to an increased injury severity when light and heavy vehicles collide (Klop and Khattak, 1999; ERSO, 2006; OECD, 2006; Kim et al., 2007; Eluru et al., 2008). In particular, road users such as pedestrians and cyclists (i.e. with no/slight mass, no/low speed and no/few protection) are more likely to be fatally injured in a road accident, especially if the collision partner rides at high speed and/or is a heavy vehicle. For instance, the probability of fatal injury for a pedestrian colliding with a motorised vehicle riding at a speed of 50 km/h is about 50-80%, whereas it reduces to 5-10% at a speed of 30 km/h (ERSO, 2006; OECD, 2006). In order to protect vulnerable road users, traffic-calming measures are often implemented in residential areas or close to specific facilities (e.g. schools). Such measures generally limit the vehicle speeds by law (e.g. 30 km/h limitation) and through road design or hurdles (e.g. loops and lollipops design, speed humps, etc.) (Pucher and Dijkstra, 2003; Pucher and Buehler, 2008; Rifaat et al., 2011), thus widening the field of vision and lengthening the perception time of motorists. Traffic-calming measures are hence expected to enhance the safety of cyclists in Brussels. Three kinds of such measures are here identified: 30 km/h, residential (20 km/h) and pedestrian areas (prohibited to motorized traffic outside delivery hours, but also to cyclists in some cases) (Appendix C.1).

4.3.3.3 Intersections (crossroads)

Intersections are known as black spots for all road users (Wang and Nihan, 2004; ERSO, 2006; Quddus, 2008; BRSI, 2009; Reynolds et al., 2009; Haque et al., 2010; Pei et al., 2010). They are places where the number of potential conflict points and the risk of having an accident are higher compared to the rest of the network (i.e. road segments) (Wang and Nihan, 2004; Geurts et al., 2005; Dumbaugh and Rae, 2009). In particular, roundabouts are often mentioned in the literature as having an unfavourable effect on cyclist safety, leading to an increased risk of accident for cyclists when they replace other types of intersections (Hels and Orozova-Bekkevold, 2007; Daniels et al., 2008; Møller and Hels, 2008; Daniels et al., 2009; Reynolds et al., 2009). This effect is even worse when the roundabout replaces a signalised intersection (compared to other types of intersections), or when marked bicycle lanes are used instead of other design types (e.g. mixed traffic or grade-separated cycle lanes) (Daniels et al., 2009). Moreover, roundabouts constructed in built-up areas and characterized by high vehicle speeds, high volumes of motorists and cyclists, multiple traffic lanes and/or large drive curves are also found to have a higher accident risk for cyclists (Hels and Orozova-Bekkevold, 2007; Daniels et al., 2008, 2009; Reynolds et al., 2009). Such findings are quite unexpected since roundabouts slow down the traffic and reduce the number of potential conflict points (compared to more conventional intersections). This also sharply contrasts with the positive safety effects observed for other road users, for whom reduced risks are generally observed (Hels and Orozova-Bekkevold, 2007; Daniels et al., 2008; Møller and Hels, 2008; Daniels et al., 2009). Besides roundabouts, signalised intersections are generally associated with reduced risks of being fatally or seriously injured when cycling (relatively to other intersections), although they may lead to an increased

risk of accident with no or slight injuries (Eluru et al., 2008; Rifaat et al., 2011). Such low levels of severity are explained by the fact that vehicle speeds and conflicting movements are reduced in these intersections (Eluru et al., 2008).
As regards the other types of intersections (e.g. right-of-ways, yield/stops, etc.), few evidence or consensus is provided in the literature about their impact on bicycle accidents. They are however expected to show higher accident risks for cyclists than 'simple' road links, since these are places where the traffic situation is more 'complex'. At such places, cyclists – as well as all road users – are faced with a large amount of information at the same time and must handle many visual stimuli (e.g. due to the dense and mixed traffic, the large number of road legs and signs, etc.) (Elvik, 2006; Dai et al., 2010). The cognitive capacity of road users is hence more likely to reach – or even exceed – its limit at intersections, which increases the probability of having an accident due to a lengthened cyclist's (or driver's) reaction time. Also, intersections with high levels of complexity (e.g. more than 4 legs with dense traffic) have high accident risks (Wang and Nihan, 2004; Elvik, 2006; Dumbaugh and Rae, 2009). In this chapter, intersections (about 10,000 in Brussels) are controlled for their evolution/infrastructure change and fall into one of the following categories: yield/stop signals, right-of-way intersection, signalized intersection (traffic lights), roundabout, intersection with right-turn lane, or pedestrian light[4] (Appendix C.1).

4.3.3.4 Tram tracks and public transport stops

Intuitively, the presence of on-road or crossable tram tracks is expected to increase the occurrence of accidents for cyclists: cyclists often declare to get one of their cycle wheels stuck in the tracks, resulting in a loss of control of their bicycle (Cameron et al., 2001; BRSI, 2006). However, no reliable evidence is provided in the literature about such a risk. Most of the research is – at our knowledge and up to now – either focussed on accidents between pedestrians and trams (see e.g. Hedelin et al., 1996; Unger et al., 2002) or indicates in a descriptive framework that the number of tram-related accidents is relatively high for cyclists, compared to other road users (Cameron et al., 2001; BRSI, 2006). Moreover, no control is made of the presence of other factors (e.g. motorised traffic, type of intersection).

The presence of public transport stops (bus, tram, metro, etc.) is also expected to cause blackspots for cyclists since frequent pedestrian activity generally occurs around these stops (Pei et al., 2010). In particular, previous studies found bus

[4] Note that pedestrian lights are not – strictly speaking – intersections, since they are generally installed in the middle of road links.

stops and bus transit intensities as being significant factors associated with the presence of bicycle accidents (Quddus, 2008; Cho et al., 2009; Pei et al., 2010). Besides the intense pedestrian activity, the poor acceleration and the large dimensions of buses probably explain to some extent such results (Walker, 2007).

Tram tracks are here digitized as linear objects on the basis of orthophotos (from BRIC, Google Earth) and using GIS data provided by the Brussels UrbIS database (BRIC) and STIB/MIVB, still over the 2006-2008 period. They are subdivided into 3 categories: tram track crossings (e.g. at crossroads), tram tracks in crossable reserved lanes (generally built parallel to the road), and on-road tram tracks (i.e. built on the road, implying that trams share the same road than cyclists and motorists). Tram tracks built in off-road separated lanes (uncrossable) are not collected, since they are separated by physical barriers or located in tunnels (they are not designed to support bicycle traffic) and hence not bikeable. As regards the public transport stops, they are digitized as point objects and are categorised into 3 classes: bus stops, tram stops, and all stops (i.e. bus, tram and metro).

4.3.3.5 Cycle facilities and discontinuities in the bicycle network

The provision of well-kept and well-planned cycle facilities is an essential ingredient for encouraging bicycle use since it reduces the actual and perceived risk associated with cycling (McClintock and Cleary, 1996; Parkin et al., 2007). When inappropriately designed and/or maintained, such facilities however carry the danger to increase the risk of cycle accidents (McClintock and Cleary, 1996). Most of the studies indeed find that cycle facilities can increase the risk of bicycle accidents compared to on-road cycling (i.e. cycling on ordinary roads, in mixed traffic) (Kaplan, 1976; McClintock and Cleary, 1996; Aultman-Hall and Hall, 1998; Aultman-Hall and Kaltenecker, 1999; Pucher et al., 1999). Although there is no consensus about the actual safety effects of each of the cycle facilities, the findings in the literature overall show that it is safer to cycle on-road than on fully segregated cycle facilities (or off-road facilities) or on cycle facilities built at intersections (Forester, 1994; Rodgers, 1997; Räsänen and Summala, 1998; Aultman-Hall and Hall, 1998; Aultman-Hall and Kaltenecker, 1999; Pucher et al., 1999; ERSO, 2006). Also, it seems that roundabouts equipped with marked cycle lanes perform significantly worse than those unequipped or equipped with other design types (Daniels et al., 2009). The lack of consensus on the results about the safety effects of the different cycle facilities probably comes from: (1) the different methodologies (more or less consistent) used to evaluate these safety effects, (2) the various definitions of cycle facilities used in the literature, (3) the way cycle facilities are designed and/or maintained in the area of interest, and (4) the spatial and temporal context in which the cycle facility is designed.

Among the possible causes that increase the risk of accident on cycle facilities, the literature often mentions the decrease of attention paid by all types of road users after the implementation of the cycle facilities (McClintock and Cleary, 1996; de Lapparent, 2005; Parkin and Meyers, 2010). In particular, the segregated facilities not only carry the risk of reducing the presence/visibility of cyclists, but also give an ill-founded feeling of safety for the cyclists. Cyclists are hence often 'unexpected' by drivers at intersections, especially when they ride in the opposite direction of the traffic (e.g. on bidirectional facilities). Such a design may indeed result in an inappropriate driver's visual search pattern (of cyclists) and may lead to the accident if the expectation of the cyclist about the driver behaviour is wrong (Räsänen and Summala, 1998). Moreover, such segregated designs increase the risk of collision with pedestrians in the case where they are shared with these latter (McClintock and Cleary, 1996). As regards on-road cycle facilities (e.g. marked lanes), Parkin and Meyers (2010) also found that drivers may give less recognition to the need to provide a comfortable passing distance when a marked cycle lane is implemented (compared to an on-road situation where there is no cycle facility).

More importantly, poorly designed facilities increase the risk of having an accident. Potential sources of danger created when building new cycle facilities may be an insufficient width of the infrastructure, an insufficient distance to the adjacent parking areas, or the creation of discontinuities or inconsistencies at some points of the cycle facility. In particular, a low width of the facility reduces the possibility to make evasive movements (e.g. in the case where there is a hurdle in the cyclist's trajectory) and increases the risk that overtaking vehicles – especially these with large dimensions, riding at high speed or passing close to the facility – throw the cyclist off his/her balance. Also, the construction of cycle facilities in the 'door zone' of parked cars may result in a potential conflict with the opening of car doors, especially if their width is insufficient and when located in built-up areas (Pai, 2011). As regards discontinuities, Krizek and Roland (2005) found they introduce high levels of discomfort when they end either on the left side of the street, on parking lots, in large intersections, or in a wider width of the curb lane.

In this chapter, both cycle facilities and discontinuities in the bicycle network are collected. On the one hand, discontinuities here correspond to the end or a cut over some distance of the cycle facility and are often observed at intersections. On the other hand, cycle facilities are defined on the basis of the terminology used by the Ministry of the Brussels-Capital Region (Brussels Mobility). They are classified into 5 categories: (1) the unidirectional separated/off-road cycle lanes, which are one-way cycle facilities located next to a road and separated by a slight elevation or any other physical barrier; (2) the bidirectional separated/off-road cycle lanes, which are two-way cycle facilities located either next to the road (with a physical separation) or fully segregated (e.g. by adopting a different trajectory compared to the road); (3) the marked cycle lanes (or bike lanes), which are one-way cycle facilities that are part of the road and marked with painted lines and/or a red-coloured surface (thus increasing the attention paid by motorists to cyclists); (4) the suggested cycle lanes (or sharrows), which are one-way cycle facilities that give cyclists (supposed safe) trajectories on the road using either different road materials or chevrons and bicycle logos; and finally, (5) the bus and bicycle lanes, which are one-way facilities dedicated to buses and cyclists. Note that suggested cycle lanes are implemented when the width is insufficient to accommodate a marked cycle lane and have the advantage to inform the motorists of the presence of cyclists in a street. They are however not subject to parking restrictions.

4.3.3.6 Parking facilities (motorised vehicles)

Parking facilities for motorised vehicles are expected to be black zones for cyclists, compared to roads without parking. Parked vehicles indeed restrict sight distances in some specific street patterns (especially when they have large dimensions) and increase the risk of conflict with exiting / parking vehicles or with car doors in the case of parallel (or longitudinal) parking facilities (Greibe, 2003; Pai, 2011; Rifaat et al., 2011). In particular, accidents due to the opening of car doors are quite frequent in urban areas since cars are here parked in great numbers along the roadside or along cycle facilities (sometimes built in the door zone) (Pai, 2011). In the densest parts of urban areas and during delivery or peak hours, vehicles are also more prone to be parked on cycle facilities, which may then force cyclists to carry out dangerous overtaking. Finally, the presence of parked vehicles after a discontinuity in the bicycle network seems to increase the level of discomfort for cyclists (Krizek and Roland, 2005). Although the number of accidents related to the presence of parking facilities is expected to be higher in urban areas, there is no evidence in the literature about what could be the actual risk of accident for cyclists riding along parked cars. Only the perceived risk of cycling is shown to be greater due to the presence of such parked vehicles along the roadside (Parkin et al., 2007).

Two types of databases are considered in this chapter: (1) 'function-based parking data', describing the role/purpose to which each parking facility is dedicated, and (2) 'aspect-based parking data', describing how parking facilities are positioned relatively to the road (e.g. in parallel or perpendicular to the road). In the first case (function-based), the facilities are subdivided into five types: (1) park-and-ride, public and private parking, (2) delivery parking, (3) diplomatic corps parking, (4) disabled parking, and (5) taxi parking. In the second case (aspect-based), the 5 following types of parking facilities are digitized at a high level of precision (i.e. accurate to within some meters) into a GIS, on the basis of the 'observed' parking behaviours (instead of considering simply the marked parking bays): (1) longitudinal parking areas, i.e. cars parked parallel to the road or the curb (on-road or on the sidewalk) and often are arranged in a line; (2) head-in (or acute) angle parking, consisting of cars parked at an acute angle with the direction of approach; (3) back-in (or reverse) angle parking, consisting of cars parked at an obtuse angle with the direction of approach; (4) parking facilities perpendicular to the road, consisting of cars parked side to side, perpendicular to the curb or road; (5) other types of parking facilities, for which there is no particular/constrained arrangement of the vehicles. Longitudinal parking areas are the most common type of parking facility in Brussels, while the other categories are often found in public places or – in some cases (as regards the parking facilities perpendicular to the road) – in residential areas. From the planner point of view, head-in angle parkings are also generally recognised as being risky for cyclists, since these latter are in the blind spot of the reversing and turning vehicles. On the contrary, back-in angle parking improve the field of vision and allow parked drivers to see passing cyclists, hence reducing the risk of collision.

4.3.3.7 Contraflow cycling

Contraflow cycling allows cyclists to travel in the opposite direction of the motorised traffic in one-way streets (Pucher et al., 2010). Contrary to popular belief, contraflow cycling is quite safe since motorists and cyclists face each other and keep a continuous eye contact (until they pass each other). It hence allows them to adapt their driving behaviour depending on the specific street features (e.g. street width, presence of longitudinal parking, etc.) and the reactions of the facing road user (Brussels Mobility). The fact that motorists generally consider contraflow cycling as unsafe also may increase the attention they pay to bicyclists while passing them in the street. This seems to be confirmed in a study conducted by Kim et al. (2007), who show that facing traffic reduces the probability of incapacitating and non-incapacitating injuries for cyclists. Contraflow cycling is hence expected to reduce the risk of accident and injury for cyclists. In this chapter, note that great care is here taken when digitising roads with contraflow cycling into a GIS. Three different data sources were used to monitor their gradual implementation (i.e. cycling maps, orthophotos, and on-line application mapping the one-way streets).

4.3.3.8 Urban facilities and public services

Partly because of the lack of data, there is little literature exploring the risk of cycling accidents associated with the proximity of specific activities or public services. To our knowledge, current research only focuses on accident frequency or severity. Disregarding the type of road user, most authors found that the number of accidents increases near employment areas, and more particularly nearby retail trade (e.g. shops, restaurants), manufacturing industry (e.g. industrial sites) and public services (e.g. schools, hospitals, etc.) (Levine et al., 1995a, 1995b; Greibe, 2003; Wedagama et al., 2006). Concerning bicycle accidents, Kim et al. (2007) also found that the presence of institutional areas (e.g. schools) increased the probability of incapacitating injury (whereas it decreased the probability of having other injury severities). In this chapter, a wide range of activities and public services are considered in order to account for the unexpected impact some of these could have on bicyclists' safety (see Appendix C.1, 'Public transport' variable).

4.3.4 Data limitations

Some data limitations are worth to mention. First, one can deplore the fact that no exposure variable (e.g. bicycle traffic flow estimation) is used. Overall, such a variable is seldom available in traffic accident research, especially for non-motorised transport modes for which less attention is generally paid by planners, scientists or policy makers (Iacono et al., 2010). As regards Brussels, the best available (exposure) data are either bicycle traffic counts performed at several locations every year (Pro Velo, 2011), or 2001 census data (FPS Economy) on the number of cyclists commuting to work or school and living in a definite statistical ward (= the smallest administrative unit in Belgium). Bicycle traffic counts are however not exploitable since they are limited over space (20 count locations only). Census data can however be used to estimate a gravity-based exposure variable (see Chapter 5), but this latter variable was not exploited here due to some technical issues in SANET in using the 'uniform network transformation' (see Okabe and Satoh, 2006). As a result of these data and technical limitations, a uniform network is here used (in the sense that the bicycle traffic is assumed to be constant over the entire road network).

Another weakness lies in the fact that no street side and/or building year of the infrastructures is taken into account, although such (detailed) data were collected for the purposes of Chapter 5. Actually, the limitations are of a methodological nature since the exploratory methods used here generalise the road links as linear features, without any precision on the street side where the infrastructure is built and/or on its building/implementation (or dismantlement) year. Such methodological issues are expected to bias the results regarding the infrastructure factors that are observed on only one street side and/or for which changes in the design are frequent (i.e. implementation of new infrastructures, changes in the street side of the infrastructure, etc.). Great care should hence be taken when analysing the results, especially as regards the network kernel density values (such densities are indeed generalised for both street sides, which might be wrong if cycling accidents occur on only one street side) and also concerning specific infrastructure variables, such as this related to streets where contraflow cycling is permitted (see Section 4.5 for a further discussion).

Reported accident data (DGSEI) and unreported ones (SHAPES) also differ on some specific points/characteristics, thus implying that the comparison between both data sets does not simply comes down to compare reported and unreported cycling accidents. The first difference is that our databases are collected using different accident registration processes (DGSEI: compulsory registration by the police; SHAPES: online registration survey). Although the electronic registration survey used for our survey allows getting more insight into the unreported minor cycling accidents (i.e. with slight injuries and/or material damage), it is however biased by the fact that it is restricted to those with access to computer and online network[5]. Second, the accident data are collected over different periods of time (DGSEI: January 1st 2006 – December 31st 2008; SHAPES: March 10th 2007 – March 16th 2009), which carries the risk that the infrastructure factors may differ from one dataset to another (for instance, for a particular street, a cycle facility might have been implemented in 2007, implying that DGSEI data do not account for its safety effect in 2006). Third, SHAPES data focus on regular[6] adult cyclists (18-65 years old) for who 60% of the cycling trips are work-related (the remaining 40% are leisure-related), whereas DGSEI data do not restrict the sample to a definite group of cyclists. This hence means that our comparative analyses will consist in comparing a sample of cycling accidents unreported by the police and involving regular and utilitarian-oriented adult cyclists (SHAPES data) with a sample of cycling accidents officially reported by the police and involving any type of cyclist (regular or not, adult or not) (DGSEI data). Lastly, as mentioned in Section 4.3.2.2, SHAPES data here involve cycling accidents that are not registered by the police and for which only small injuries (i.e. bruises or cramps) and/or material damages are reported, whereas cycling accidents reported by the police – and then by the DGSEI – seem to have a

[5] The access (and use) of a computer and internet is still nowadays strongly associated with the age. In Belgium, about 60% of the households had an access to internet in 2007 (which rose to 73% in 2010) (FPS Economy).

[6] Regular cyclists are here defined as cyclists commuting at least twice a week to their workplace.

slightly higher degree of injury severity (i.e. body injuries with either short- or long-term consequences for the cyclist).

Last but not least, it is noteworthy that SHAPES data (unreported accidents) present some limitations. Besides the bias caused by the online data collection (see above), one can also deplore the fact that serious injuries and fatalities are expected to be strongly underreported in the survey. Accident-related data may indeed be not encoded anymore in the case where a serious fatality (physical disability) or a fatality occurs. The small number (49) of observations collected within the SHAPES online survey is also an important limitation, as it may affect the significance of the results. This is even more problematic when focussing on the Pentagon and the First Crown, which are characterised by smaller extents/areas and smaller sample sizes (9 observations are reported in the Pentagon, while 34 are observed in the First Crown). Confidence envelopes of the expected values computed for the network (cross) K-functions are however larger in such cases (low number of observations), which involves that the null hypothesis for CSR is less ‘easily’ rejected. Note that Figure 4.3 illustrates well such a statement.

4.4 Methodology

4.4.1 Comparative statistics and odds ratios

In a first step, comparative statistics are computed in order to identify whether or not (un-) reported accidents are more likely to be associated with specific factors/variables (see Section 4.3.3 for further description). Such statistics here consist of Chi-Square adjusted tests and Fisher’s exact tests for independence (as regards discrete data), as well as Wilcoxon Rank-Sum tests (continuous data). The Chi-Square adjusted test – which is a continuity-adjusted version of the Pearson Chi-Square test (i.e. adjusted for the continuity of the Chi-Square distribution) – and the Fisher’s exact test are both used for discrete/nominal factors, characterized by small sample sizes. They test whether there is a significant difference between unreported and reported accidents in terms of spatial factors. The Wilcoxon Rank-Sum test (or Mann-Whitney U test) is a non-parametric alternative to the two-sample Student’s t-test[7] and is here used in the case where factors are measured on a continuous scale. It tests whether these latter significantly differ (in their median values) between unreported and reported accidents.

[7] In most cases, the data assumptions of normality are not valid.

In a second step, odds ratios (*OR*s) and their lower and upper confidence intervals are computed in order to compare the odds of observing a specific factor in the unreported accident data set compared to the odds of observing it in the reported accident data set. In other words, they give us an insight of how likely a specific factor is observed at unreported accident locations, compared to reported ones. As a corollary, *OR*s might hence be helpful in identifying the locations where cycling accidents are the most likely to be unreported.

4.4.2 Point pattern analyses in traffic-accident research

Although spatial analysis of road accidents generally relies on data aggregated over definite spatial units and time periods, there are also some studies regarding each individual accident as a single point in space (with coordinates x,y) and aiming at exploring and/or understanding the spatial distribution of these points over a specific period of time (see e.g. Levine et al., 1995a; Jones et al., 1996; Yamada and Thill, 2004; Myint, 2008). On the one hand, *segment- or area-based analyses* are often conducted for administrative convenience, time constraints or in the case where accident data are available in aggregate form only (e.g. counts per road link or area). Such analyses have the advantage to eliminate some of the year-to-year fluctuations in the individual location of accidents, but they have the drawback to produce spatial errors (in the sense that accidents are not anymore individual locations in space) and lead to results that are dependent upon the set of spatial units on which the data are aggregated (Nicholson, 1985; Bailey and Gatrell, 1995; Levine et al., 1995b; Lawson, 2009). This latter problem – often referred in the literature to as the 'modifiable areal unit problem' (MAUP) – means that the modification of the size of the units is likely to conduct to different results and conclusions. On the other hand, *point pattern analyses* may be adopted whether data are available at the individual accident level, i.e. in the case where accurate information is available about the location of accidents in space (Yamada and Thill, 2004). Overall, most methods implemented for point pattern analyses either measure the global variation in the mean value of the spatial process (*first-order effects*), or examine the tendency for local deviations from the mean value caused by the spatial correlation structure of this process (*second-order effects*) (Bailey and Gatrell, 1995; O'Sullivan and Unwin, 2002). Methods investigating first-order effects are e.g. quadrat count analyses and kernel density estimations, while second-order effects are measured using e.g. nearest-neighbour distances and K-functions (Cressie, 1993; Bailey and Gatrell, 1995; Fotheringham et al., 2000).

Point pattern analyses are here conducted for exploring (and comparing) the spatial patterns of reported and unreported cycling accidents. First, centrographic methods and kernel density estimations (in planar and network spaces) are briefly described in Section 4.4.2.1, and afterwards used as methods for initial point pattern exploration. Second, the basic concepts of the K-function and cross K-function methods are presented for the planar space as well as for the network space (Section 4.4.2.2). K-function methods are helpful in depicting the spatial distribution of both reported and unreported accident data sets, while cross-K function methods are used to examine the spatial distribution of these accidents with respect to specific (spatial) factors.

4.4.2.1 Initial point pattern exploration

Centrographic methods

Centrographic methods consist of measures of central tendency and spatial dispersion of the spatial point pattern. Four measures are here used: (1) spatial mean centre, (2) central feature, (3) standard distance, and (4) standard deviational ellipse. First, the spatial mean centre identifies the average location of the point pattern, i.e. the mean latitude and mean longitude of all the point events; it hence corresponds to the centre of gravity of this point pattern. Second, the central feature provides the most centrally located feature in a spatial point pattern. Third, the standard distance (or standard distance deviation) measures the standard deviation of the point pattern around the mean centre, i.e. the degree of spatial dispersion or compactness of the point distribution around this centre. Last but not least, the standard deviational ellipse computes the directional trend of a point distribution. This latter method calculates the standard deviation separately for the x and y coordinates (from the mean centre), which then defines the axes of the so-called standard deviational ellipse. The major axis of the ellipse is in the direction of maximum dispersion and is at right angles to the minor axis (which is in the direction of minimum dispersion). In other words, such a measure then exhibits the spatial dispersion and the direction/orientation of a point distribution in space.

Although these measures are useful in summarizing a point distribution, they have the drawback to be affected by outliers and do not investigate the second-order effects of the distribution (i.e. the spatial interactions between points). For further details on these measures, refer to Ebdon (1985), Fotheringham et al. (2000), Myint (2008) and ESRI (2009).

Kernel density estimations (KDE)

Kernel density estimation (KDE) is commonly used to estimate the density of points in space (Bailey and Gatrell, 1995). Such a technique computes a smooth estimate of a probability density over space from an observed point pattern. Visually, it may evoke three-dimensional humps (or kernels) placed at locations **s** and then summed over space to obtain a density estimate for the point distribution (Cressie, 1993; Bailey and Gatrell, 1995; Fotheringham et al., 2000). Formally, the density (or intensity) at location $\boldsymbol{s}$ is noted $\lambda(\boldsymbol{s})$ and is defined as:

$$\lambda(\boldsymbol{s}) = \sum_{i=1}^{n} \frac{1}{h^2} k\left(\frac{\boldsymbol{s} - \boldsymbol{s}_i}{h}\right) \tag{4.1}$$

where $\boldsymbol{s}$ is a location in the studied area $\Re$ (or 'study region'), $\boldsymbol{s}_1$, ..., $\boldsymbol{s}_n$ are the locations i of the n observations/events (e.g. cycling accidents; $i = 1, ..., n$), $k()$ is the kernel function, and h is the bandwidth (also called the smoothing parameter or window width). The bandwidth corresponds to the radius of a circle centred on $\boldsymbol{s}$ and within which observations $\boldsymbol{s}_i$ are 'taken into account' to compute $\lambda(\boldsymbol{s})$. Its selection determines the amount of smoothing of the data: large bandwidths will exhibit flat densities $\lambda(\boldsymbol{s})$ and will highlight regional patterns, whereas small bandwidths will lead to spiky densities (centred on $\boldsymbol{s}_i$) and will underscore local patterns (Bailey and Gatrell, 1995; Fotheringham et al., 2000). The kernel function $k()$ is a probability density function used to determine the distance decay effect within the bandwidth (Bailey and Gatrell, 1995; Xie and Yan, 2008). In the literature, the most commonly used kernel functions are of Gaussian, Quartic, Minimum variance, Epanechnikov, negative exponential, or Conic functional forms (Fotheringham et al., 2000; Schabenberger and Gotway, 2005; Xie and Yan, 2008). The kernel function and − more particularly − the bandwidth are hence two key parameters about which the analyst has to make choices (see Silverman (1986) and Brunsdon (1995) for a further discussion on the selection of these parameters).

In the case where the spatial phenomenon is analysed on a network (such as road accidents), the KDE as defined in Equation 4.1 is likely to provide biased estimates since it assumes that the study region is represented by a homogeneous two-dimensional planar space, where the distances are Euclidean (Yamada and Thill, 2004; Okabe et al., 2006a, 2006b; Xie and Yan, 2008; Okabe et al., 2009). A KDE based on network distances between point events would indeed be more appropriate (relative to Euclidean distances) when these events occur only on a one-dimensional subset of the planar space (i.e. the network). For instance, a planar KDE applied to a network-constrained distribution of points could lead to high densities detected at locations $\boldsymbol{s}_i$, whereas lower densities could be obtained when applying the KDE on a network space (imagine e.g. two close parallel streets, without any intersection and with one accident on each of these). As a consequence, the planar assumption applied to the distribution of network-constrained points is no longer valid. Okabe et al. (2009) recently extended the ordinary KDE method (or 'planar KDE') to a network space, assuming that: (1) point events are constrained on a network, and (2) distances between two of these points are computed on that network (instead of being Euclidean-based). In this chapter, a network KDE called the '*equal split discontinuous kernel function*[8]' is estimated from SANET in order to get insight about the density of (reported) cycling accidents on the Brussels' network. Interestingly, this kernel function satisfies five properties: it is unbiased, unimodal, symmetric with respect to two kernel centres, invariant with respect to a vertex angle, and the kernel centre coincide with the modal point. Nevertheless, the estimator does not satisfy continuity at each node of the network, as well as it may lead to unequal densities for equal distances in the kernel (in the case where vertices are present) (see Okabe et al. (2009) for further details).

4.4.2.2 Univariate and bivariate *K*-function analyses

Planar and network K-functions (univariate analysis)

The reduced second moment measure or Ripley's *K*-function (Ripley, 1976, 1981) is commonly used for analysing the spatial distribution of observed events/points over a wide range of scales on an infinite homogeneous plane (Cressie, 1993; Bailey and Gatrell, 1995; Jones et al., 1996; Fotheringham et al., 2000). Assuming a planar space, the *K*-function – noted $K(h)$ – is defined as follows:

$$K(h) = \frac{1}{\lambda} E\begin{pmatrix} \text{number of points in a set } P \text{ within Euclidean} \\ \text{distance } h \text{ of a randomly chosen point in } P \end{pmatrix} \tag{4.2}$$

where λ is the intensity of the studied point process P (i.e. the number of points in a given set P divided by the area of the study region), $E[.]$ is the expectation operator, and distance $h \geq 0$. More concretely, the *K*-function for a given Euclidean distance h corresponds to the average number of points counted in a circle of radius h around a (randomly chosen) point in P, divided by the intensity of the point process P. In particular, a suitable estimate of $K(h)$ for an

[8] Note that a 'continuous' function also exists. This latter makes the kernel function continuous around the vertices of the network. It is not recommended in the case where the network includes many short links, since it increases the computational complexity of the function (Okabe et al., 2009).

observed set of n points (p_1, ..., p_n) distributed over a study region with area R is given by (Diggle, 1983; Boots and Getis, 1988; Gatrell et al., 1996):

$$\hat{K}(h) = \frac{1}{\lambda n} \sum_{i=1}^{n} \sum_{i \neq j}^{n} I_h(d_{ij}) = \frac{R}{n^2} \sum_{i=1}^{n} \sum_{i \neq j}^{n} I_h(d_{ij}) \tag{4.3}$$

where $\lambda = n/R$, d_{ij} is the Euclidean distance between the points p_i and p_j, and $I_h(d_{ij})$ is an indicator function which is 1 if $d_{ij} \leq h$, and 0 otherwise. In order to test whether the observed point distribution is regular, clustered or random, the K-function estimated for the observed distribution – i.e. $\hat{K}(h)$ – is compared with the theoretical value of $K(h)$ obtained under complete spatial randomness (CSR), following the homogeneous Poisson point process (see Cressie (1993) for further information on the CSR concept). The null hypothesis for CSR is hence tested in order to detect the presence of clustering or dispersion in the observed point pattern. If the points are uniformly and independently distributed over space (i.e. under CSR), the expected number of points within an Euclidean distance h of a randomly chosen point is $\lambda\pi h^2$, and then the theoretical value $K(h) = \pi h^2$ for all values of h. This suggests that the observed points are spatially clustering if $\hat{K}(h) > \pi h^2$, while $\hat{K}(h) < \pi h^2$ indicates the presence of regularity in the observed point pattern (i.e. the points are repelling or are dispersing over space). The null hypothesis for CSR is then rejected if the observed points are spatially clustering or repelling, i.e. whether there is a significant deviation of the observed estimate $\hat{K}(h)$ from a randomly generated point process (estimated by $K(h) = \pi h^2$).

Besides the fact it allows spatial dependence to be analysed over a wide range of scales, the K-function has the advantage to handle *all* point-to-point Euclidean distances to analyse the point distribution on a planar space, whereas the nearest neighbour analysis just accounts for the *nearest* neighbor distances between points. However, as for KDE (in Section 4.4.2.1), the assumption of a continuous infinite plane is problematic in the case where the point process is inherently constrained on a (finite) network space (Okabe and Yamada, 2001). For such a network-constrained process, the use of the K-function over a planar space – which is termed here the 'planar K-function' – would indeed result in the over-detection of clustered patterns (so leading to possible Type I errors) since the actual network distances between points are underestimated when computed over a planar space (Yamada and Thill, 2004; Dai et al., 2010). This hence suggests that the planar K-function should be extended to a network space. For a set of points P distributed according to the binomial point process along a finite network L_T, Okabe and Yamada (2001) then define the network K-function as:

$$K^{net}(h) = \frac{1}{\rho} E\begin{pmatrix} \text{number of points in } P \text{ within network} \\ \text{distance } h \text{ of a point } p \text{ on } L_T \end{pmatrix},\ h \geq 0 \qquad (4.4)$$

where $\rho = n/|L_T|$ is the number of points of P divided by the total network distance $|L_T|$ (i.e. the density of points over L_T), and $E(.)$ is the expected value with respect to all possible locations of p, that are assumed to follow a stochastic point process called the 'binomial point process'. The assumption of the binomial point process is based on the hypothesis that all points of P are uniformly and independently distributed over the network L_T (in other words, all points are located at random over L_T). This hence suggests that points of P are spatially interacting if this hypothesis is rejected (e.g. they may spatially cluster or repel). For an observed point pattern $(p_1, \ldots, p_n)$, a suitable estimate of $K^{net}(h)$ is given by (Okabe and Yamada, 2001; Yamada and Thill, 2004):

$$\hat{K}^{net}(h) = \frac{|L_T|}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j}^{n} I_h(s_{ij}) \qquad (4.5)$$

where s_{ij} is the network distance between the points p_i and p_j, and $I_h(s_{ij})$ is an indicator function which is 1 if $s_{ij} \leq h$, and 0 otherwise. Equation 4.5 is here called the 'observed network K-function' for a definite set P of points[9]. As limitation, such a formulation disregards the point distribution outside the study region and hence does not correct the first type of edge effect[10]. It however has the advantage to eliminate the second type of edge effect since the statistic (in Equation 4.5) is properly extended to a finite space (i.e. the network space) rather than being based on the strong assumption that the (network-constrained) point process occurs on an infinite planar space (Okabe and Yamada, 2001; Yamada and Thill, 2004; Myint, 2008). It is hence not necessary to adjust the formulation of the network K-function as it is commonly done in the planar case (Okabe et al., 2006b).

For both planar and network K-functions, confidence envelopes (or intervals) of the expected/theoretical values (i.e. $K(h)$ or $K^{net}(h)$) are estimated under the null hypothesis of CSR (binomial point process) in order to test the randomness of the observed point pattern at all possible scales. In other words, regarding the network case in particular, the statistical test consists in comparing the observed values of the network K-function $\hat{K}^{net}(h)$ (computed using the observed data)

[9] See Okabe and Yamada (2001) and Yamada and Thill (2004) for further details about the formulation of the network K-function.

[10] Two types of edge / boundary effects exist in spatial statistics. It occurs (1) when points outside the study region are disregarded, and (2) when a statistic that is based on the assumption that a point process occurs in an infinite space is applied to a finite space (Okabe and Yamada, 2001; Yamada and Thill, 2004; Myint, 2008).

to the envelope of the expected/theoretical values of the network *K*-function $K^{net}(h)$ computed under CSR. Monte Carlo simulations are here used to compute (under CSR) the expected values of $K^{net}(h)$ as well as their upper and lower significance intervals at the 5% (pseudo-) significance level. Then, if the values of $\hat{K}^{net}(h)$ lie within the confidence envelope of $K^{net}(h)$ at a definite distance, we may conclude that the randomness of the observed point pattern is not rejected at that distance. If $\hat{K}^{net}(h)$ is above the upper interval of $K^{net}(h)$, then it suggests that the observed points are spatially clustering and that the randomness of the observed point pattern may be rejected. Conversely, if $\hat{K}^{net}(h)$ is below the lower envelope, then the randomness of the observed point pattern may be rejected and the distribution of points tends towards significant regularity or dispersion, i.e. they are repelling over space (Bailey and Gatrell, 1995; Spooner et al., 2004; Yamada and Thill, 2004; Deckers et al., 2005). Within the framework of this chapter, the univariate network *K*-function analyses are performed using SANET (Okabe et al., 2006a, 2006b), with the aim to test whether reported and unreported cycling accidents tend to cluster (or repel) over a network space.

Planar and network cross K-functions (bivariate analysis)

The 'cross-*K* function method' – also called the 'bivariate *K*-function method' – is used to compare the distribution of two sets of points, *A* and *B*. Such a method allows examining whether the points in *A* tend to cluster, disperse or locate at random with respect to the points in *B* (Cressie, 1993; Bailey and Gatrell, 1995). To examine such (spatial) relationships between *A* and *B*, we make the null hypothesis that points in *A* are distributed according to a homogeneous Poisson point process (i.e. under CSR). This assumption implies that points in *A* are uniformly and independently distributed over space, regardless of the distribution of points in *B* (note that no assumption is made with respect to this latter). Considering the planar case, the above hypothesis is examined by defining the cross *K*-function of *A* relative to *B* (Bailey and Gatrell, 1995):

$$K_{ba}(h) = \frac{1}{\lambda_a} E\begin{pmatrix} \text{number of points in } A \text{ within Euclidean distance} \\ h \text{ of a randomly chosen point } b \text{ in } B \end{pmatrix},\ h \geq 0 \qquad (4.6)$$

where λ_a is the density of points of *A* ($\lambda_a = n_a/R$, where n_a is the total number of points in the set *A*), and $E(.)$ is the expected value of the number of points in *A* (which follow a homogeneous Poisson point process) with respect to the points in *B*. Assuming two observed point patterns A ('non-basic' points) and B ('basic' points), the observed cross-*K* function of *A* relative to *B* is estimated as follows (*ibid.*):

$$\hat{K}_{ba}(h) = \frac{R}{n_a . n_b} \sum_{i=1}^{n_b} \sum_{j=1}^{n_a} I_h(d_{ij}) = \frac{1}{\lambda_a . n_b} \sum_{i=1}^{n_b} \sum_{j=1}^{n_a} I_h(d_{ij}) \tag{4.7}$$

where $A = \{ a_{j=1}, \ldots, a_{n_a} \}$ and $B = \{ b_{i=1}, \ldots, b_{n_b} \}$, n_a is the total number of points a_j in A, n_b is the total number of points b_i in B, d_{ij} is the distance between a_j and b_i, and $I_h(d_{ij})$ is an indicator function which is 1 if $d_{ij} \leq h$, and 0 otherwise. In order to test the null hypothesis, the observed values $\hat{K}_{ba}(h)$ are then compared to the expected / theoretical values $K_{ba}(h) = \pi h^2$ obtained under CSR. If $\hat{K}_{ba}(h)$ significantly deviate from $K_{ba}(h)$, then the null hypothesis is rejected and it may be inferred that points A (Bailey and Gatrell, 1995):

(1) either tend to spatially cluster around points B, if $\hat{K}_{ba}(h)$ > upper intervals of $K_{ba}(h)$

(2) or tend to spatially repel around points B, if $\hat{K}_{ba}(h)$ < lower intervals of $K_{ba}(h)$

In other words, the cross-K function method aims at examining the locational tendency of non-basic points A with respect to basic points B, i.e. if points A spatially cluster, repel or distribute at random around points B (Myint, 2008).

Regarding the network case, biased results and conclusions may be obtained if the planar cross-K function is used (instead of its network equivalent) to analyse the spatial interactions between two inherently network-constrained sets of points (Okabe et al., 2006a). As a result, Okabe and Yamada (2001) then extended the formulation of the planar cross-K function to a network space. They define the network cross-K function of A relative to B as:

$$K_{ba}^{net}(h) = \frac{1}{\rho_a} E \begin{pmatrix} \text{number of points in } A \text{ within network} \\ \text{distance } h \text{ of a point } b \text{ in } B \end{pmatrix}, \; h \geq 0 \tag{4.8}$$

where ρ_a is the density of points of A on the network ($\rho_a = n_a/|L_T|$) and $E(.)$ is the expected value of the number of points in A (which follow a binomial point process) with respect to the points in B. Considering two observed point patterns $A = \{ a_{j=1}, \ldots, a_{n_a} \}$ and $B = \{ b_{i=1}, \ldots, b_{n_b} \}$ that are constrained to occur on a network, the observed network cross-K function of A relative to B is given by (Okabe and Yamada, 2001):

$$\hat{K}_{ba}^{net}(h) = \frac{|L_T|}{n_a . n_b} \sum_{i=1}^{n_b} \sum_{j=1}^{n_a} I_h(s_{ij}) = \frac{1}{\lambda_a . n_b} \sum_{i=1}^{n_b} \sum_{j=1}^{n_a} I_h(s_{ij}) \tag{4.9}$$

where s_{ij} is the network distance between the points a_j and b_i, and $I_h(s_{ij})$ is an indicator function which is 1 if $s_{ij} \leq h$, and 0 otherwise. Once again, the null hypothesis is tested by comparing the observed values $\hat{K}_{ba}^{net}(h)$ with the expected values $K_{ba}^{net}(h) = \pi h^2$ obtained under CSR (according to the binomial point process). Monte Carlo simulations are used to estimate the expected values and their confidence envelopes at the 5% significance level. In this chapter, the observed and expected values of the network cross-K functions are estimated using SANET, in order to inspect whether the reported and unreported cycling accidents significantly cluster (or repel) with respect to definite spatial factors.

4.5 Results and discussion

Empirical analyses are here conducted to explore/compare the spatial patterns and locational tendencies (around specific infrastructure factors) of reported and unreported cycling accidents. Section 4.5.1 first presents the results of comparative statistics and odds ratios in order to get a first insight into the relationships between the reported and unreported cycling accidents (in terms of the observed infrastructure factors). Such a preliminary step is then completed by an initial point pattern analysis (Section 4.5.2), aiming at exploring the spatial distribution of these cycling accidents through the implementation of centrographic methods and KDE. Section 4.5.3 then ends with the results of the univariate and bivariate network K-functions, which aim at examining if reported and unreported cycling accidents both cluster over space and/or if they concentrate around specific infrastructure factors.

4.5.1 Comparative statistics and odds ratios

Tables 4.1 and 4.2 exhibit a few statistics aiming at comparing the infrastructure factors observed for reported (DGSEI) and unreported (SHAPES) cycling accidents. Depending on the type of variable (discrete or continuous), different comparative statistics are provided in those tables. Discrete factors (Table 4.1) represent factors for which the presence/absence of a specific infrastructure is noted 1/0, while continuous factors (Table 4.2) correspond to network-based distance measures between the cycling accidents and the infrastructures under study. In Table 4.1, both Chi-Square adjusted tests and Fisher's exact tests for independence indicate that – in most cases – the type of accident (unreported / reported) is not significantly associated with a particular type of infrastructure. This hence suggests that the reported and unreported cycling accidents overall occur at places characterized by similar infrastructure factors. Odds ratios (OR) not only confirm these results (overall, OR values are around 1), but also quantify the odds of observing an unreported cycling accident for a definite infrastructure factor, compared to reported accidents.

Significant associations (accident–infrastructure) are however highlighted in Table 4.1. In comparison with reported cycling accidents, our findings show that unreported accidents are about 3 times more likely to occur in areas where 30km/h speed limits are imposed and, more generally, in traffic-calming areas. The reduced differential between the speed of slow and fast modes (created by the lower speed limits) probably explains such a result. In the case where they collide with a motorized vehicle in such areas, the cyclists generally incur slighter injuries (and/or material damages) and do not feel the need to call the police. A high rate of underreporting then results from such a lower degree of severity of the accidents (so explaining why unreported cycling accidents seem to be more likely to occur in traffic-calming areas). Regarding the places where no cycle facility is built as well as the streets where contraflow cycling is permitted, our results also suggest that cycling accidents are more likely to be unreported here than elsewhere. Such findings should nevertheless be interpreted with great caution since the survey data (SHAPES) do not include any information about the traffic direction of the cyclist involved in the (unreported) accident. This remark is even more true with respect to the cycling accidents occurring in streets where contraflow cycling is permitted[11]. In such a case, a bias is expected to occur and may lead to a wrong interpretation of the results.

In Table 4.2 (continuous factors), the Wilcoxon Rank-Sum tests – also called 'Mann-Whitney tests' – suggest that there are significant differences in the proximity to specific locations along the network (i.e. facilities, services, etc.) between reported and unreported cycling accidents. Interestingly, the unreported accidents seem to occur closer (compared to the reported ones) to European administrative buildings, superior schools (i.e. high schools and universities), shopping centres, cultural buildings, hospitals, and specific types of parking areas (park-and-ride, taxi, public and private parking areas). As for traffic-calming areas, the greater occurrence of unreported cycling accidents close to most of these facilities and services is expected to be explained by a lower differential of speed between cyclists and motorized vehicles. Except for parking areas, several types of traffic-calming measures are generally taken in the neighbourhood of such attractive facilities / services[12] in order to reduce the risk and the severity of accidents involving vulnerable road users (such as pedestrians and cyclists).

[11] In such a case, the unreported cycling accidents are (erroneously) assumed to occur in the contraflow direction if the accident occurred in a street where contraflow cycling is allowed.

[12] The term 'attractive' here refers to the fact a consistent number of trips is attracted by the facility / service.

For instance, speed limits (30km/h) and physical measures (e.g. speed humps) are frequently implemented near schools and hospitals, while pedestrian areas are often observed in the proximity of shopping centres and cultural buildings (mainly located in the historic centre of the city, in the case of Brussels). Such measures are then expected to reduce the degree of accident severity and – as a result – the registration rate among (slight) cycling accidents.

As a result of these comparative statistics, it can be concluded here that reported and unreported cycle accidents exhibit similar locational tendencies, i.e. they distribute in a similar way around specific types of road infrastructures and facilities. Areas where there is a lower differential of speed between fast and slow road users however constitute an exception to this general rule, thus suggesting that cycling accidents are more likely to be unreported here. Great care should then be taken when analysing (reported) cycling accidents in these areas.

4.5.2 Initial point pattern exploration

Figure 4.2 illustrates the point distributions of unreported and reported cycling accidents (respectively), as well as the centrographic measures computed for these. As expected, these maps show that most of reported and unreported cycling accidents occur in the Pentagon and in the First Crown, i.e. in districts where the densities and the number of cycling trips (generated and/or attracted by these high densities) are high. Strikingly, they also indicate that the spatial mean centres computed for each distribution are very close to each other, and that the central features locate at the same place in Brussels (i.e. in the Central Business District, near the European and regional administrations). Regarding the standard distances, the spatial distribution of reported cycling accidents exhibits the highest deviation from the spatial mean, whereas unreported accidents tend to be less spatially dispersed. The standard deviational ellipses finally provide further information in highlighting a northwest-southeast orientation for both spatial distributions of accidents. As a conclusion of these four graphical measures, we clearly suggest that unreported cycling accidents distribute over space in an analogous way to reported cycling accidents (and vice versa), all the more so the lower spatial dispersion of unreported cycling accidents is probably explained by the low number of accidents collected through the on-line survey (especially for suburbs located in the Second Crown).

Table 4.1: Infrastructure factors (discrete) – Descriptive and comparative statistics

	Ψ (description)	N_S	(%)	N_D	(%)	χ^2 test (p)	F test (p)	OR	(LCI-UCI)
Bridge	-	0	(0.0)	12	**(2.0)**	0.65	1.00	n.a.	n.a.
Tunnel[†]	-	0	(0.0)	0	(0.0)	n.a.	n.a.	n.a.	n.a.
Traffic-calming area	1 (30 km/h)	12	**(24.5)**	55	(9.2)	**0.00**	**0.00**	3.21	(1.58-6.52)
	2[†] (pedestrian)	0	(0.0)	3	**(0.5)**	1.00	1.00	n.a.	n.a.
	3[†] (residential)	0	(0.0)	2	**(0.3)**	1.00	1.00	n.a.	n.a.
	4 (all types)	12	**(24.5)**	60	(10.0)	**0.00**	**0.01**	2.92	(1.44-5.90)
Crossroad	0 (no crossroad)	17	(34.7)	266	**(44.3)**	0.25	0.23	0.67	(0.36-1.23)
	1 (yield/stop)	6	**(12.2)**	66	(11.0)	0.98	0.81	1.13	(0.46-2.75)
	2 (right-of-way)	12	**(24.5)**	111	(18.5)	0.40	0.34	1.43	(0.72-2.83)
	3 (traffic light)	11	**(22.4)**	106	(17.7)	0.52	0.44	1.35	(0.67-2.73)
	4 (roundabout)	3	(6.1)	40	**(6.7)**	1.00	1.00	0.91	(0.27-3.07)
	5[†] (right-turn)	0	(0.0)	9	**(1.5)**	0.82	1.00	n.a.	n.a.
	6[†] (pedestrian light)	0	(0.0)	2	**(0.3)**	1.00	1.00	n.a.	n.a.
Tram tracks[a]	0 (no tram track)	40	(81.6)	495	**(82.5)**	1.00	0.85	0.94	(0.44-2.00)
	1 (crossing tracks)	1	(2.0)	34	**(5.7)**	0.45	0.51	0.35	(0.05-2.59)
	2 (reserved lanes)	3	**(6.1)**	22	(3.7)	0.64	0.43	1.71	(0.49-5.94)
	3 (on-road tracks)	5	**(10.2)**	49	(8.2)	0.82	0.59	1.28	(0.48-3.37)
Cycle facility[a]	0 (no cycle facility)	46	**(93.9)**	486	(81.0)	**0.04**	**0.02**	3.60	(1.10-11.77)
	1 (unidirectional)	2	(4.1)	30	**(5.0)**	1.00	1.00	0.81	(0.19-3.49)

continued on next page

continued

	Ψ (description)	N_S	(%)	N_D	(%)	χ^2 test (p)	F test (p)	OR	(LCI-UCI)
Cycle facility[a]	2 (bidirectional)	0	(0.0)	22	**(3.7)**	0.34	0.40	n.a.	n.a.
	3 (marked lane)	1	(2.0)	42	**(7.0)**	0.30	0.24	0.28	(0.04-2.06)
	4 (suggested lane)	0	(0.0)	15	**(2.5)**	0.53	0.62	n.a.	n.a.
	5† (bus/bicycle lane)	0	(0.0)	5	**(0.8)**	1.00	1.00	n.a.	n.a.
Parking area (aspect-based)[a]	0 (no parking area)	25	(51.0)	348	**(58.0)**	0.42	0.37	0.75	(0.42-1.35)
	1 (longitudinal)	22	**(44.9)**	245	(40.8)	0.69	0.65	1.18	(0.66-2.12)
	2† (head-in angle)	0	(0.0)	2	**(0.3)**	1.00	1.00	n.a.	n.a.
	3† (back-in angle)	0	(0.0)	1	**(0.2)**	1.00	1.00	n.a.	n.a.
	4† (perpendicular)	1	**(2.0)**	2	(0.3)	0.55	0.21	6.23	(0.55-69.94)
	5† (other types)	1	**(2.0)**	2	(0.3)	0.55	0.21	6.23	(0.55-69.94)
Contraflow cycling[a]	-	15	**(30.6)**	32	(5.3)	**0.00**	**0.00**	7.83	(3.87-15.84)

[a] Variables for which DGSEI and SHAPES accidents are not entirely comparable, given that the street side where the infrastructure is built (or where the measure comes into effect) is not taken into account for SHAPES accidents

† Less than 10 observations for both SHAPES and DGSEI accidents; care must be taken when analyzing the corresponding data

Ψ: Nominal variable, taking on different values for each infrastructure variable (one value = one kind of infrastructure or facility; see Appendix C.1 for further details)

n.a.: not available (insufficient number of observations / accidents)

N_S (%): number and percentage (%) of SHAPES accidents (**bold**: % SHAPES accidents > % DGSEI accidents)

N_D (%): number and percentage (%) of DGSEI accidents (**bold**: % DGSEI accidents > % SHAPES accidents)

χ^2 test (p): p-value of the Chi-Square adjusted test for independence (**bold**: independence not rejected)

F test (p): p-value of the Fisher's exact test for independence (**bold**: independence not rejected)

OR: Odds Ratio; LCI: Lower Credible Interval of the OR (2.5%); UCI: Upper credible interval of the OR (97.5%)

Table 4.2: Infrastructure factors (continuous) – Descriptive and comparative statistics

	Ψ (description)	SHAPES accidents		DGSEI accidents		Wilcoxon test (p)
		D_{mean}	D_{std}	D_{mean}	D_{std}	
Discontinuity	-	325.0	281.5	356.4	337.2	0.74
Parking area (function-based)	1 (park&ride, public, private)	467.1	292.6	629.5	461.6	**0.02**
	2 (delivery)	311.1	314.3	407.0	473.2	0.18
	3 (diplomatic corps)	662.1	555.0	915.0	908.2	0.17
	4 (disabled)	172.7	112.2	205.3	252.3	0.82
	5 (taxi)	396.4	239.2	639.4	590.4	**0.01**
	6 (all types)	111.4	95.3	142.6	191.0	0.51
Public transport	1 (bus stop)	390.0	328.0	360.9	327.2	0.38
	2 (tram stop)	635.7	445.0	683.3	602.2	0.90
	3 (all types of stops)	345.8	314.0	283.1	266.8	0.14
Public administration	1 (European buildings)	1550.3	1213.7	2170.2	1617.6	**0.01**
	2 (regional buildings)	1657.9	1291.8	1774.8	1311.2	0.49
	3 (all types of buildings)	1129.4	972.7	1458.8	1180.5	0.07
School	1 (primary or secondary)	376.2	222.8	389.6	265.6	0.88
	2 (international prim./sec.)	1437.2	1053.3	1884.2	1456.3	0.06
	3 (superior)	687.2	574.2	938.7	820.9	**0.02**
	4 (all types)	266.4	167.1	335.3	250.3	0.06

continued on next page

continued

	Ψ (description)	SHAPES accidents		DGSEI accidents		Wilcoxon test (p)
		D_{mean}	D_{std}	D_{mean}	D_{std}	
Industrial estate	-	1897.5	731.7	1780.6	955.8	0.11
Shopping center	-	1290.9	1113.5	1723.1	1297.8	**0.01**
Supermarket	-	649.4	424.1	754.4	629.7	0.47
Service station	-	473.4	239.9	539.7	334.3	0.41
Cultural building	-	435.2	313.5	611.3	516.3	**0.02**
Sports complex	-	1125.9	550.3	1119.5	547.8	0.87
Playground	-	653.3	361.7	618.8	373.3	0.44
Religious building	1 (synagogue)	2386.9	1612.5	2775.5	1825.9	0.19
	2 (protestant)	763.3	521.7	2775.5	1825.9	**0.00**
	3 (orthodox)	1553.1	1199.0	1767.8	1388.5	0.37
	4 (mosque)	1110.8	842.6	1416.3	1200.8	0.22
	5 (catholic)	486.9	245.2	530.5	319.1	0.59
	6 (all types)	406.1	248.0	411.5	310.0	0.78
Police building	-	865.7	535.1	850.3	520.1	0.84
Hospital	-	983.5	745.8	1197.9	850.7	**0.03**
Embassy	-	722.0	569.6	1031.0	973.4	0.10

Ψ: Nominal variable, taking on different values for each infrastructure variable (one value = one kind of infrastructure or facility; see Appendix C.1 for further details)

D_{mean}: average network distance to the closest 'point feature' (e.g. public transport stop, discontinuity) (in meters)

D_{std}: standard deviation of network distances (accidents-closest point features)

Wilcoxon test (p): p-value of Wilcoxon Rank-Sum test (Mann-Whitney). Significant differences are in **bold.**

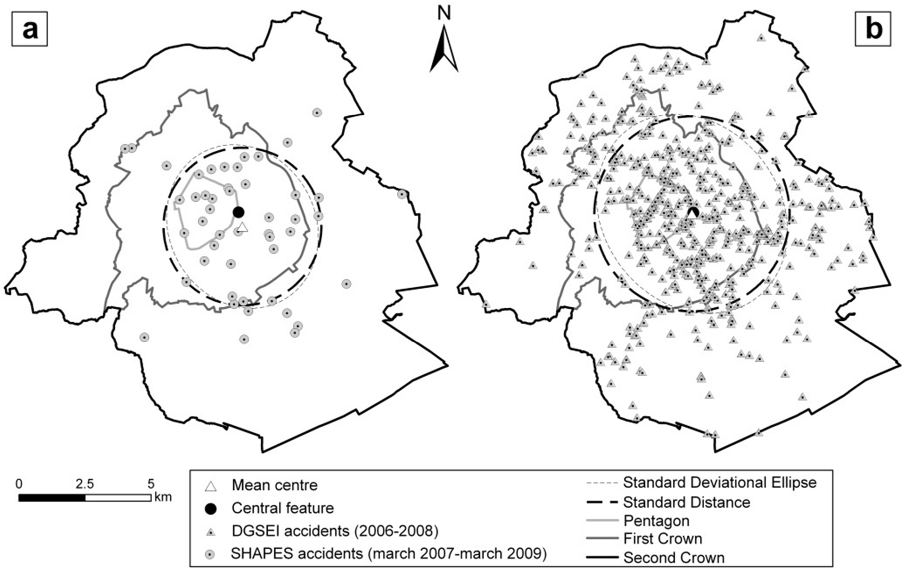


Figure 4.2: Centrographic measures for the distribution of (a) unreported cycling accidents (SHAPES survey) and (b) reported cycling accidents (DGSEI data)

Network kernel densities are also computed with the aim to: (1) get a first insight into the location of black spots of cycling accidents along the network; and (2) visually identify the infrastructure factors that could play a role in the occurrence of cycling accidents, at the scale of the BCR. The equal-split discontinuous kernel method is here applied in SANET v.4 (beta) to compute the densities of cycling accidents. As illustration, Appendix C.2 zooms in the Brussels' Pentagon and shows the network densities in the case where cycling accidents are officially reported by the police (DGSEI data)[1]. Such exploratory results prove to be useful in identifying (visually) the factors that could play a role in the occurrence of the (reported) cycling accidents. Unsurprisingly, segments with high densities of accidents are observed at major intersections (i.e. intersections made up of a large number of road legs) as well as on roads characterized by busy traffic conditions and passing through dense employment areas (e.g. near to shopping centres). Examples of such high-density segments are the boulevards oriented in a southwest-northeast direction in the Pentagon (referred as 'A' in Appendix C.2) and the intersections between the inner pentagon-shaped ring road and the major avenues (B). These results are in line

[1] As regards unreported cycling accidents, the densities are not illustrated here because of the small sample size collected for the Pentagon (which provides an incomplete representation of the black spots of cycling accidents).

with previous results in the literature (see e.g. Anderson, 2009). Conversely, low-density segments (< 0.2) are mainly observed along roads with low volumes of motorised traffic and going across residential districts (C). Regarding the densities obtained for other parts of the BCR (in the First and Second Crowns), it also turns out that cycling accidents cluster at discontinuities along the bicycle network and on roads equipped with on-road tram tracks (and – to a lesser extent – on roads equipped with crossable reserved lanes).

4.5.3 Network *K*-functions and cross *K*-functions

Network *K*-function and cross *K*-function methods are carried out at the scale of the three spatial subareas (i.e. the Pentagon, the First Crown and the Second Crown) in order to examine if the results differ from one subarea to another. Both methods are conducted in SANET v.4 beta[2] and use 500 Monte Carlo simulations to estimate the expected network *K*-function as well as the 95% upper and lower confidence intervals.

Table 4.3: Analysis of the spatial distribution of both unreported (SHAPES) and reported (DGSEI) cycling accidents, at the scale of 3 different subareas

		Network *K*-function		**Network cross *K*-function**†	
Study region	**Database (n)**	**Pattern**	**d_c (m)**	**Pattern**	**d_c (m)**
Pentagon	DGSEI (82)	N	$d_c < 750$	N	$d_c < 200$
	SHAPES (9)‡	N	$d_c = \varnothing$		
1st Crown	DGSEI (356)	C	$d_c \geq 0$	C	$d_c \geq 0$
	SHAPES (34)	C	$d_c > 120$		
2nd Crown	DGSEI (600)	C	$d_c \geq 0$	C	$d_c \geq 0$
	SHAPES (49)	C	$d_c > 120$		

† Basic points: DGSEI data; Non-basic points: SHAPES data
‡ Small number of observations; great care is hence required when analyzing the results
n: number of points
C: spatial clustering; N: no spatial pattern (randomness or independence)
d_c: distances values where significant spatial clustering is observed

[2] Note that the SANET team reports as minor error that the constant of the cross-*K* function is not divided by the number of basic points in SANET v4.beta. This however does not affect the statistical test of spatial randomness and the interpretation of the results (see http://sanet.csis.u-tokyo.ac.jp/sub_en/errata.html).

On the one hand, the univariate network *K*-function method is used to test whether or not the reported (DGSEI) and unreported (SHAPES) cycling accidents cluster, repel or distribute at random along the network. Table 4.3 indicates that, for the First and Second Crowns, both reported and unreported cycling accidents significantly cluster at almost all values of network distance (unreported accidents spatially cluster beyond 120m). As illustration for the Second Crown, graphics on the left side of Figure 4.3 indeed show that the observed values of the *K*-function (grey line) are to the left of the 5% upper confidence interval (upper dashed black line). In contrast, unreported accidents in the Pentagon are randomly distributed at all distances (the grey line appears within the 95% envelope in Figure 4.3, down right), while reported accidents only cluster up to a 750m distance and then distribute at random for larger distances (Figure 4.3, up right).

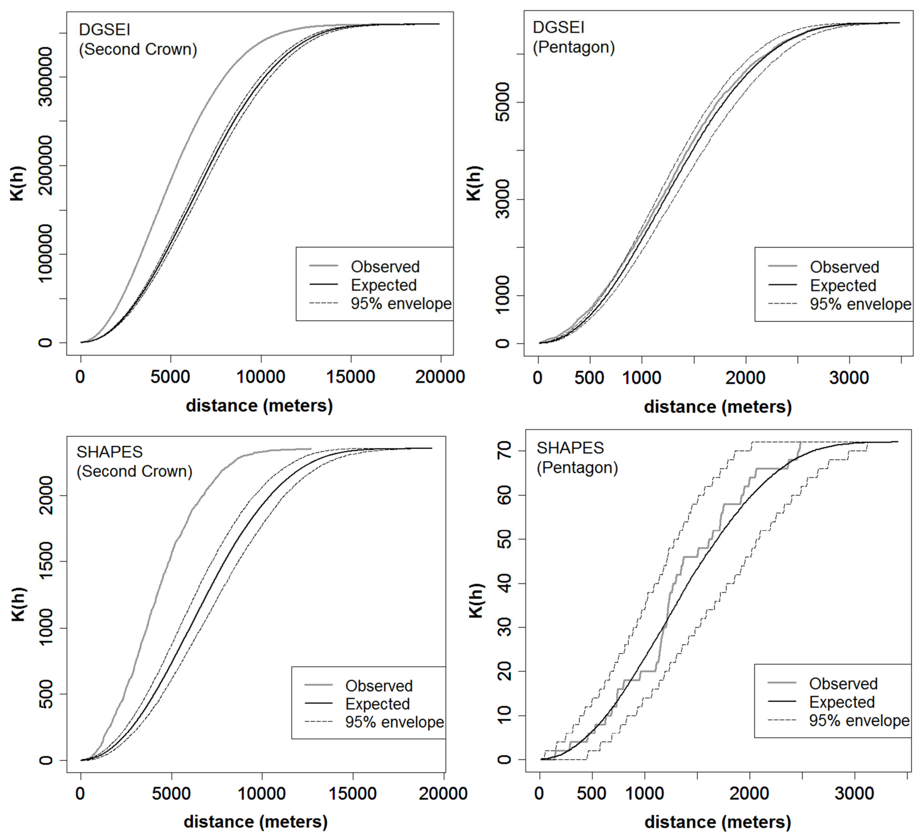


Figure 4.3: Univariate spatial pattern analysis of both unreported (SHAPES) and reported (DGSEI) cycling accidents – Network *K*-function, Brussels' Pentagon and Second Crown

The conclusions are then twofold: (1) the inferences about an observed spatial point pattern (e.g. spatial clustering, randomness, regularity) may differ depending on the spatial subarea considered by the analyst; (2) for a given spatial subarea, reported and unreported cycling accidents tend to distribute in a same/close way along the network (e.g. they *both* cluster along the network for a definite subarea).

On the other hand, the bivariate network cross *K*-function method is applied in order to examine whether unreported and reported cycling accidents are observed in the vicinity of each other, and whether they have similar or different locational tendencies with respect to specific infrastructure factors. Table 4.3 and Figure 4.4 indicate that unreported cycling accidents tend to be located around reported accidents at the scale of the First and Second Crowns, whereas this only turns out to be the case for the shortest distances (< 200m)[3] in the case of the Brussels' Pentagon. Such findings hence support the fact that the results – and the interpretation of these latter – strongly depend on the chosen spatial subarea on which the point pattern analyses are conducted. They also bring some additional pieces of evidence that unreported and reported cycling accidents locate in the vicinity of each other, which suggests that they could occur at places characterized by similar infrastructure factors. This is confirmed in Table 4.4, where unreported and reported cycling accidents rarely show dissimilar point patterns when distributing around a definite infrastructure factor at the scale of a given spatial subarea. This is even truer as regards the Pentagon and the Second Crown, for which there is not the slightest dissimilarity in the patterns (probably because the Pentagon and the Second Crown are 'too' small and 'too' large spatial subareas, respectively). The selection of a definite spatial subarea may then be of importance, which suggests that there could be one subarea more suitable than another. Within the framework of this chapter, the First Crown is probably the best compromise, although it is here thought that several subareas may provide complementary information. For instance, they may be used as a helpful mean to check the consistency of the results, to select an appropriate spatial subarea, or to detect at which scale and from which distance threshold an observed point pattern (e.g. the cycling accidents) spatially clusters around specific locations (e.g. unreported cycling accidents gather around industrial estates beyond 5400m and only in the case where the Second Crown is the chosen spatial subarea).

When observed, the dissimilarities (in the spatial point patterns) consistently occur at the scale of the First Crown and mostly concern bridges, marked cycle

[3] Note that network distances between 450 and 800m also show significant clustering over the network, in the case where reported cycling accidents are used as non-basic points in SANET (Table 4.3, Figure 4.4 on the left side).

lanes, head-in angle parking areas, and industrial estates. Although not significant for the shortest distances, spatial clustering of reported cycling accidents is observed around such locations, whereas the spatial pattern tends to be random as regards the unreported accidents. Except for industrial estates and both religious buildings, such a result is probably explained by the fact that the injury severity – and, then, the registration of the accidents – is higher when the cycling accident occurs on a marked cycle lane (door-related accidents), a bridge (reduced space), or near head-in angle parking (blind spot accidents). This could in turn explain why unreported cycling accidents distribute at random around these locations. Another explanation could be the low number of accidents registered by the SHAPES survey, although it might not be the most plausible one since the results are overall comparable to these obtained from the DGSEI data (whatever the spatial subarea).

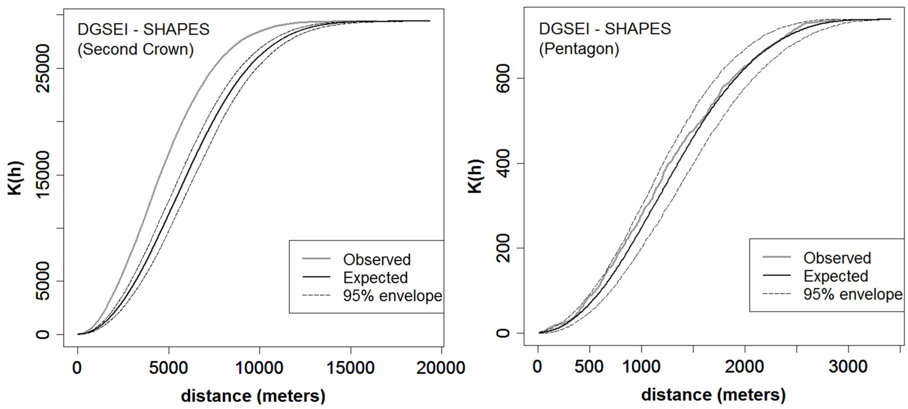


Figure 4.4: Locational tendency of unreported cycling accidents (SHAPES) with respect to reported cycling accidents (DGSEI) – Bivariate spatial pattern analysis, using the network cross *K*-function, Brussels' Pentagon and Second Crown

In line with the previous results focussing on the Pentagon, Table 4.4 also indicates that both unreported and reported cycling accidents invariably distribute at random with respect to each infrastructure factor. Conversely, cycling accidents tend to gather around most of these factors (at all values of distances) in the case where the First or Second Crown is selected as spatial subarea. The presence of crossroads, tram tracks, discontinuities in the bicycle network, schools, shopping centres, or parking areas – among other factors – generally tend to be spatially associated with (unreported and reported) cycling accidents.

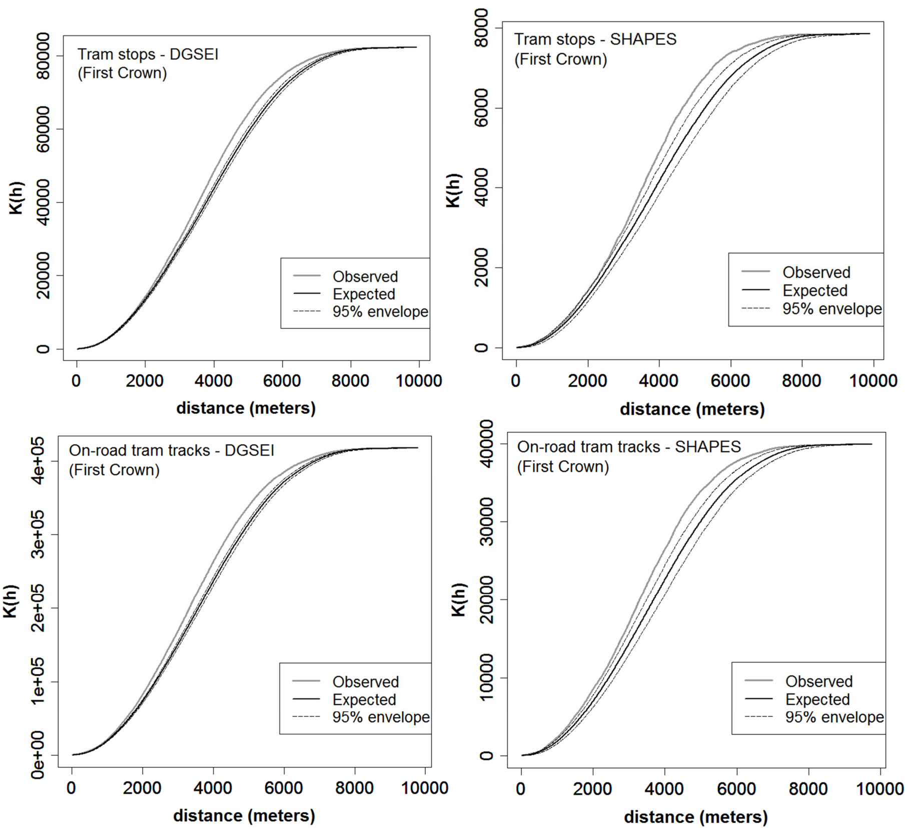


Figure 4.5: Locational tendency of both unreported (SHAPES) and reported (DGSEI) cycling accidents with respect to: (1) tram stops (up), and (2) on-road tram tracks (down) – Bivariate spatial pattern analysis, using the network cross K-function and carried out at the scale of the First Crown

Although little dissimilarity in the *overall* spatial patterns is noted between unreported and reported cycling accidents (especially as regards the Pentagon and the Second Crown), some subtle differences can however be emphasized at some (short) ranges of distances and/or in the level of significance of the spatial clustering of accidents around infrastructures. Such differences are not only present between unreported and reported accidents (the reported ones showing the highest levels of significance), but are also noted between the infrastructure factors. For instance, our results show that – at the scale of the First Crown – unreported cycling accidents significantly cluster around tram stops beyond about 2000m of network distance (Figure 4.5, up right), whereas reported cycling accidents show significant spatial clustering for all values of distance (Figure 4.5, upper left). As a comparison, both reported and unreported cycling accidents more significantly cluster around on-road tram tracks than around tram stops

(Figure 4.5, down) given that the observed curve is to the left of the expected one for almost all values of network distance (except for a short range of distance spreading from 370 to 670m, as regards unreported accidents).

Table 4.4: Analysis of the spatial distribution of reported and unreported cycling accidents, with respect to infrastructure factors and using 3 spatial subareas (Pentagon, 1st and 2nd Crowns)

Infrastructures (basic points)	Ψ (description)	n_P	n_{1C}	n_{2C}	P	1C	2C
Bridge[a]	-	7	85	159	NN	**CN**	FF
Tunnel[a]	-	10	74	156	NN	NN	FF
Traffic-calming area[a]	1 (30 km/h)	213	837	2589	NN	CC	FF
	2 (pedestrian)	142	176	198	NN	CC	FF
	3 (residential)	6	17	81	NN	CC	FF
	4 (all types)	361	1030	2868	NN	CC	FF
Crossroad	1 (no crossroad)	122	711	1660	NN	CC	FF
	2 (yield/stop)	417	2587	6061	NN	CC	FF
	3 (right-of-way)	147	671	1193	NN	CC	FF
	4 (traffic light)	38	418	1263	NN	CC	FF
	5 (roundabout)	45	235	568	NN	CC	FF
	6 (pedestrian light)	12	51	143	NN	CC	FF
Tram tracks[a,b]	1 (crossing tracks)	74	649	1147	NN	CC	FF
	2 (reserved lanes)	26	227	341	NN	CC	FF
	3 (on-road tracks)	31	313	604	NN	CC	FF
Cycle facility[a,b]	1 (unidirectional)	6	137	605	NN	NN	FF
	2 (bidirectional)	6	78	480	NN	NN	FF
	3 (marked lane)	53	404	899	NN	**CN**	FF
	4 (suggested lane)	23	133	193	NN	CC	FF
	5 (bus/bicycle lane)	42	106	118	NN	CC	FF
Parking area (aspect-based)[a,b]	1 (longitudinal)	1430	9802	21196	NN	CC	FF
	2 (head-in angle)	34	302	700	NN	**CN**	FF
	3 (back-in angle)	2	32	92	NN	NN	FF
	4 (perpendicular)	96	501	1437	NN	CC	FF
Contraflow cycling[a,b]	-	480	2034	3375	NN	CC	FF
Discontinuity	-	71	385	684	NN	CC	FF
Parking area (function-based)	1 (park & ride, public, private)	29	75	156	NN	CC	CC
	2 (delivery)	191	575	737	NN	CC	FF
	3 (diplomatic corps)	18	242	384	NN	CC	CC
	4 (disabled)	136	1296	2268	NN	CC	FF
	5 (taxi)	24	93	136	NN	CC	CC
	6 (all types)	398	2281	3681	NN	CC	FF

continued on next page

continued

Infrastructures (basic points)	Ψ (description)	n_P	n_{1C}	n_{2C}	P	1C	2C
Public transport	1 (bus stops)	25	525	1050	NN	CC	FF
	2 (tram stops)	18	300	505	NN	CC	CC
	3 (all types)	43	793	1485	NN	CC	CC
Public administration	1 (European buildings)	0	53	66	NN	CC	FF
	2 (regional buildings)	10	27	30	NN	CC	FF
	3 (all types of buildings)	10	80	96	NN	CC	FF
School	1 (primary, secondary)	42	269	574	NN	CC	CC
	2 (international)	0	11	22	NN	CC	CC
	3 (superior)	22	55	85	NN	CC	FF
	4 (all types)	64	335	681	NN	CC	CC
Industrial estate	-	1	11	33	NN	**RN**	CC
Shopping center	-	13	23	28	NN	CC	FF
Supermarket	-	6	53	110	NN	CC	CC
Service station	-	5	77	194	NN	CC	FF
Cultural building	-	63	143	200	NN	CC	CC
Sports complex	-	5	17	57	NN	NN	FF
Playground	-	10	60	187	NN	CC	FF
Religious building	1 (synagogue)	2	9	12	NN	NN	FF
	2 (protestant)	15	104	130	NN	**CN**	FF
	3 (orthodox)	2	16	18	NN	CC	FF
	4 (mosque)	5	68	74	NN	**CN**	CC
	5 (catholic)	13	64	127	NN	CC	FF
	6 (all types)	37	261	361	NN	CC	CC
Police building	-	4	28	53	NN	CC	FF
Hospital	-	5	17	38	NN	CC	FF
Embassy	-	8	103	184	NN	CC	CC

[a] Linear objects/features

[b] The street side (where the infrastructure is built) is not taken into account

Ψ: Nominal variable, taking on different values for each infrastructure variable (one value = one kind of infrastructure or facility; see Appendix C.1 for further details)

C: spatial clustering; N: no spatial pattern (randomness or independence); R: regularity (or dispersion); F: failed to compute or lack of time. The first letter refers to DGSEI accidents, and the second one to SHAPES accidents (e.g. CN = spatial clustering for DGSEI accidents, and no spatial pattern for SHAPES accidents)

n_P, n_{1C}, n_{2C}: number of points in the Pentagon, First Crown and Second Crown (respectively)

-: not applicable

4.6 Conclusions

The objective of this exploratory chapter was to provide further knowledge about the spatial distribution of 'unreported' cycling accidents, i.e. those that are not officially registered by the police and the DGSEI. It aimed at analysing whether or not unreported and reported cycling accidents showed similar spatial patterns on a network space (e.g. if they cluster with respect to each other along this network) and if they both had the same locational tendencies with respect to specific road infrastructures (e.g. if they both spatially cluster around bridges). Focussing on the Brussels-Capital Region, this chapter took advantage of combining official DGSEI data (= cycling accidents reported by the police, involving mainly slight body injuries with short- or long-term consequences for the cyclist) with those collected through an open-based online registration survey and for which there is no police record (= unreported cycling accidents, resulting in bruises, cramps and/or material damages).

Comparative statistics and spatial point pattern analyses – using a combination of centrographic, KDE and network *K*-function methods – have shown to be useful in exploring (and comparing) the spatial distributions of reported and unreported cycling accidents. Comparative statistics first reveal that both reported and unreported cycling accidents tend to occur at rather similar locations, i.e. at locations where similar road infrastructures and activities are observed. More interestingly, they also suggest that cycling accidents are more prone to be unreported by police in areas where there is a lower differential of speed between cyclists and motorized vehicles (e.g. in traffic-calming areas, where speed limits and physical measures are frequently implemented by planners in order to lower the speed of motorists). Such a lower differential indeed reduces the injury severity of cycling accidents and – as a corollary – decreases the need to call the police (given that the cyclist can cure oneself and/or repair oneself the material damages). It then implies that registration efforts should be concentrated on areas where traffic-calming measures are taken (e.g. in the vicinity of schools, 30km/h areas, pedestrian areas, residential areas, etc.), especially if the purpose is to improve the recording of cycling accidents. Great caution is also recommended when analysing official databases of road accidents in these areas, given that underreporting rates of (slight) cycling accidents are expected to be higher here than anywhere else. Ignoring this may clearly lead to a biased interpretation about the safety effects related to traffic-calming strategies (especially as regards the accident severity).

Centrographic and network (cross) *K*-function methods support our previous results in indicating that unreported and reported cycling accidents show similar spatial patterns along the network and both cluster around the same infrastructures. Although the definition of both types of accidents is not perfectly equivalent, our findings hence suggest that improving the accident registration for cyclists (e.g. through surveys, like ours) would not necessarily provide further knowledge about (unobserved) spatial factors associated with the occurrence of cycling accidents (except for a few factors, such as head-in parking areas). Official accident databases may then serve as a good basis for orienting policy decisions and (safety-oriented) investments at a regional scale, although a more complete registration of cycling accidents is required (and even recommended) if local safety treatments are intended by planners and/or policy-makers. Conversely, it also suggests that our survey data (SHAPES) may be considered as spatially representative of official accident databases and then hold the potential to provide some good insights in the actual spatial patterns of reported cycle accidents. Lastly, our findings not only highlight strong similarities in the locational tendencies of reported and unreported cycling accidents, but also emphasize the importance to select an appropriate spatial subarea for conducting point pattern analyses. It is indeed demonstrated here that the results of the network (cross) K-function methods strongly depend on the chosen spatial subarea and – hence – that they should be interpreted with great caution. At best, these methods should be carried out on several spatial subareas.

From a methodological point of view, one can however deplore some major limitations here. K-function methods indeed have the drawback to be unable to test the significance of clustering along the network (Yamada and Thill, 2004). Also, cross-K function methods do not account for potential interrelationships between the infrastructure factors. For instance, the locational tendency of cycling accidents to cluster around one specific infrastructure factor does not necessarily mean that a causal relationship exists between this factor and cycling accidents. Such a tendency may be entirely explained by the presence of another (correlated) factor that plays a more prominent role in the occurrence of accidents than the infrastructure factor with which it is correlated. It is hence tricky to draw here reliable conclusions on the separate safety effect related to each specific infrastructure factor. As a consequence, policy recommendations are to be strongly avoided within the framework of this point pattern exploration. A multivariate framework would then be of great help to control for the presence of other (correlated) factors and to estimate the importance of such separate effects. This is however beyond the scope of this chapter, which is here exploited as an initial exploratory data analysis before moving on a modelling step (which is approached in the following chapter).

Chapter 5

Accident risk when cycling in Brussels
An innovative spatial case-control approach[1]

Outline

Bicycle use provides an effective way of addressing health, environmental and mobility concerns in urban areas. However, accident risks strongly deter people from cycling. Identifying the factors having an impact on such a risk is helpful in coping with the numerous fears and safety concerns inhabitants have about bicycling. This chapter then aims at understanding the spatial distribution of bicycle accidents in Brussels (Belgium) with the intent to provide safety-oriented policy recommendations. A spatial Bayesian modelling approach is here proposed to model the spatial variation of accident risks for cyclists (2006-2008 period), using a binary dependent variable (accident, no accident at location *i*) constructed from an innovative case-control strategy. Control sites are sampled along the 'bikeable' road network and as a function of the potential bicycle traffic transiting/stopping in each Brussels' statistical ward. Risk factors are either infrastructure-related (e.g. type of intersection), traffic-related (e.g. van and truck traffic) or environmental (e.g. topography). Our findings suggest that a higher risk of accident is statistically associated with the presence of on-road tram tracks, bridges (without any cycle facility), complex intersections, close shopping centres, garages, and higher volumes of van and truck traffic. Cycle facilities built at intersections (especially suggested cycle lanes at right-of-way intersections) and parked vehicles located next to separated cycle facilities (i.e. in the 'door zone') also increase this risk, whereas streets where contraflow cycling is permitted reduce it (outside intersections). More interestingly, mapping the

[1] This chapter will be submitted in 2011 for publication.

predicted accident risk along the network provides for planners and policy makers a value-added tool that accurately locates the places at high risk of accident and where cycling accidents might have been unreported.

5.1 Introduction

As pointed out before, the spatial point pattern methods used for network analysis in Chapter 4 do not fully account for the potential interrelationships existing between the factors (expected to be) associated with an increased frequency of cycling accidents. This chapter hence extends the exploratory data analyses conducted in the previous chapter by carrying out the statistical analyses within a *multivariate framework* and accounting for multicollinearity, which allows estimating the individual/separate effects of the (explanatory) factors while controlling for the possible correlations between these latter. Contrary to the previous chapters (for which a so-called 'frequentist' approach is chosen), a Bayesian computational approach is here preferred as it provides several advantages over the estimation carried out in a frequentist framework. The ability to incorporate prior expert knowledge and to deal with nuisance/random parameters (i.e. unobserved heterogeneity) in complex models is one of the key assets of the Bayesian approach (Koop, 2003; Miaou et al., 2003; Bolstad, 2007; Kéry, 2010). Unlike frequentist inference that generally gives point (or fixed) estimations, the Bayesian approach allows the parameters to be characterised as random variables and provides direct probability statements about these[2] (Bolstad, 2007; Kéry, 2010; Pei et al., 2010). Probability is hence expressed as the uncertainty we have about the magnitude of a parameter, which makes the Bayesian inference more intuitive compared with the conventional approaches (for which the probability is the relative frequency of a feature observed in our data set). Frequentist inference may also be biased when using finite sample sizes, whereas Bayesian computational methods give exact inference for any sample size (Kéry, 2010). The advent of Markov Chain Monte Carlo (MCMC) methods as well as the availability of softwares that implement such simulation-based approaches (e.g. WinBUGS, MLwiN) are at the root of the growth in popularity of Bayesian methods. This, combined with a continuous improvement of the computer technologies (e.g. improved storage, computer processing speed, etc.), made most of the complex Bayesian models computationally tractable. MCMC methods – such as Metropolis Hastings or

[2] In Bayesian statistics, probability statements are made about a parameter, rather than about a data set (as it is the case in frequentist statistics). This hence means that popular statements such as 'I am 99% sure it will rain tomorrow' can only be derived from a Bayesian framework, whereas they are not valid using frequentist statistics (Kéry, 2010).

Gibbs sampling algorithms – are techniques that iteratively draw samples from the so-called 'posterior distribution' of a parameter. Gibbs sampling (which is used in WinBUGS) is a special case of the Metropolis Hastings algorithm and has the advantage to handle complex problems like simple ones, i.e. the complex problem is broken down into smaller units, which are then solved one at a time (Clark, 2005; Gelman and Hill, 2007; Lawson, 2009; Kéry, 2010).

Compared to Chapter 4, this part of the thesis also adds some improvements by integrating the *street side* and the *building year* of the infrastructures/facilities (as far as possible). More importantly, the background *exposure* of cyclists to accidents is also taken into account here. Chapter 4 indeed highlighted the need to account for the exposure of cyclists to road accidents, given that cycling accidents are expected to occur in greater numbers in places where the bicycle traffic is high (and conversely). It is hence of utmost importance to account for such a traffic variable if the purpose is to properly evaluate the risk of running a cycling accident at specific places or provide some recommendations about the safety effects associated with specific types of infrastructures or facilities.

Unlike most of the previous research aiming at developing models of road accident severity or frequency, the main objective of this chapter is hence to explain the *risk* of having an accident for a cyclist on the entire road network of the Brussels-Capital Region (urban area), using spatial risk factors as covariates of a statistical model as well as a gravity-based approach in order to account for the exposure of cyclists in the traffic. Such an estimation of the accident risk on an *entire* road network – rather than considering only *specific* locations or parts of the network (e.g. road trajectories selected by the analyst, which is likely to orient/bias the statistical results; see e.g. Lusk et al. (2011) and attendant comments) – merits further consideration since planners and policy makers are generally interested to know what are the most significant infrastructure-related factors (and then the locations) associated with high accident risks (whatever the mode of transport). This chapter attempts to provide an answer to such a major concern, starting from accident data only (i.e. without any available controls at the beginning of the analysis) and focussing on bicycle accidents. As a result, the specific aims of this chapter are the following: (1) identifying which are the most significant spatial variables/factors (expected to be) associated with the occurrence of a bicycle accident in an urbanised area (i.e. Brussels), (2) identifying which areas are expected to carry the highest risk or probability to generate bicycle accidents (based on model predictions), and (3) providing recommendations intended for policy makers and planners (Chapter 6). The methodology is innovative in the sense that the modelling framework uses an autologistic model combining accident data (from police) *and* control points (i.e. exposure of cyclists) in order to predict the *risk* of having an accident (rather than the severity of the accident). Also note that the scale of the analysis is the accident itself and that special attention is paid to spatial autocorrelation during the Bayesian modelling process (besides multicollinearity).

This chapter is organised as follows. Section 5.2 defines preliminary concepts, describes the models within the Bayesian framework and motivates the use of a case-control approach. Section 5.3 describes the data used in this chapter. Finally, Section 5.4 reports the main results, after which Section 5.5 concludes this chapter. Note that the results we obtained here also serve as basis for some of the scientific-based recommendations provided in Chapter 6.

5.2 Conceptual and methodological framework

The lack of detailed data on accidents and trip characteristics associated with the different modes of transport often hamper researchers to improve their understanding on the factors affecting the probability of accidents (Lord and Mannering, 2010). Depending on the size of the studied area and the level of aggregation, collecting data on accident-related mechanisms (e.g. road user behaviour at the moment of the accident), risk factors (e.g. type of parking areas) or exposure data (e.g. traffic flow estimation) can be cost- and time-consuming, especially if they are not available or when working at local scales of analyses. As a result, most of the statistical models aggregate the accidents and their risk factors over spatial units (at various scales, e.g. at the scale of road segments, municipalities, counties, regions or even countries) and/or over a definite time period (Aguero-Valverde and Jovanis, 2006; Liu and Jarrett, 2008; Quddus, 2008; Lord and Mannering, 2010).

Overall, traffic accident research either aims at predicting the frequency of accidents, or attempts to explain the association between various severity or collision types and several independent variables (Noland and Quddus, 2004; Lord and Mannering, 2010). Some of these models however may lead to several well-known methodological issues (e.g. over- or under-dispersion of accident-frequency data, low sample means and size, injury severity and accident-type correlation, etc.) and hence require performing appropriate statistical approaches in order to avoid incorrect inferences that could result from these data-related problems (Lord and Mannering, 2010).

Instead of routinely modelling either the accident severity or the accident frequency, we here implement a case-control methodology aiming at modelling the accident risk for a cyclist along an entire road network (i.e. with both road intersections and sections). Except when both the accident data and the trip patterns (with the exact trajectories of the road users) are available (through e.g. a detailed survey; see e.g. Harris et al., 2011), such an estimation of the accident risk presupposes the generation of controls, i.e. the creation of data reflecting the exposure of the population under study (i.e. the cyclists) to the outcome of interest (i.e. the accident). Once generated, such controls can be coupled with the accident database in order to produce a binary dependent variable as well as to make possible the estimation of accident risks for cyclists through the use of logistic regressions (performed within a Bayesian framework). However, a great care must be taken when generating such controls since they are likely to bias the results if they are not selected within a rigorous statistical framework and if no control is made of some important risk factors. In the literature, some studies already attempted to estimate the risks of having an accident, but did not select the location of their controls – i.e. their exposure data – in a rigorous way (e.g. they are often selected in a town centre and on frequent trajectories of the road users of interest, without any well-founded statistical basis), or did not completely account for the spatial variability of some important risk factors in their analysis (e.g. variability in traffic volume, variability in terms of the types of intersections and parking areas, etc.) (see e.g. Lusk et al., 2011, and the attendant comments). As a consequence, results are likely to be biased and might lead to wrong conclusions about some risk factors.

Let us first define some basic concepts used in this chapter (Section 5.2.1), before reviewing some of the main concepts and modelling approaches from which one's inspiration was drawn to generate the controls in a rigorous way (Section 5.2.2). The description of the so-called '*accident-risk models*' implemented in this chapter is finally approached in Sections 5.2.3 and 5.2.4.

5.2.1 Pre-requisites

As mentioned in Chapter 1, a ***bicycle accident*** refers to any road accident involving at least one cyclist. In this chapter, it is defined regardless of the trip purpose, accident severity, age and gender. Further details on the construction of the dependent variable – i.e. having a bicycle accident or not – are provided in the section dedicated to the data collection.

The notion of ***risk*** is the probability that the outcome of interest (i.e. the bicycle accident) will occur, following a particular exposure of the population or study group (Burt, 2001; Porta, 2008). Within the framework of this chapter, the risk of having an accident for a cyclist is the probability that this accident will occur, following the exposure of the cyclists in the traffic during a specified period of time (2006-2008, in this case). This exposure is here constructed/defined on the basis of a potential (or gravity-based) measure of the bicycle traffic in Brussels, which is expected to be proportional to the levels of cycling observed at different parts/locations of the region (see Section 5.2.3.1).

Risk factors – or risk indicators – refer to 'independent' variables (also called covariates, explanatory variables, etc.) that affect the probability of a specified outcome of interest, such as the occurrence of a bicycle accident. Risks factors in traffic accident research are not necessarily causal factors (e.g. they can contribute to the occurrence of the accident only if they are combined *with* other risk factors), and some of these can be modified by intervention(s) aiming at reducing the probability of accident (e.g. infrastructure change, modification of the behaviours, etc.). The notion of 'modifiable risk factors' is then logically used in this last case (*ibid.*).

5.2.2 From ecology and epidemiology…

The methodological framework implemented in this chapter mainly draws one's inspiration from the research in epidemiology and ecology, taking advantage of their respective methodological strengths in modelling and case-control strategies. On the basis of an extensive review of the literature, it turns out that a lot of work has been done in order to develop techniques addressing the issue of the lack/absence of controls in databases (especially in ecological modelling). The purpose of such techniques is to provide some information on the absence of the outcome of interest (i.e. the disease or the observation of a species), which in turn enables to pair presences and absences in a same database in order to use common regression techniques based on binary data (e.g. logistic regression). Such approaches will be replicated in this chapter with the aim to obtain binary data (presence-absence) and compute the risk of a bicycle accident using logistic regression modelling. Approaches in ecology are first briefly reviewed, after which we focus on case-control studies used in epidemiology.

5.2.2.1 Presence-only data

In ecology, most of the available data on species consist of so-called 'presence-only' data sets, i.e. where data on locational records (i.e. observation or collection of a species at a particular location) are available with some degree of accuracy – depending on the quality of the ground surveying –, but for which there is no information on the absence of species (Ferrier et al., 2002; Zaniewski et al., 2002). Two groups of techniques are generally used for modelling the distribution of a species using such presence-only data: (1) the profile techniques, i.e. those incorporating the presence-only data into the model (e.g. environmental envelopes, genetic algorithms, or ecological niche factor analyses (ENFA)), and (2) the group discrimination techniques, i.e. those requiring the generation of 'pseudo-absence points' ('pseudo' because the probability of having a 'true absence' is not absolutely certain) in order to supplement the presence-only data and hence facilitate the use of logistic regression modelling (Brotons et al., 2004; Engler et al., 2004; Guisan et al., 2007; Zarnetske, 2007; Chefaoui and Lobo, 2008; Wisz and Guisan, 2009). Overall, the second group of techniques is often preferred to profile techniques since it is derived from well-established statistical approaches and provides more accurate predictions (Chefaoui and Lobo, 2008; Wisz and Guisan, 2009). Group discrimination techniques either (1) generate the pseudo-absences at random over space, or (2) weight the random sampling of the pseudo-absences in favour of areas expected to contain 'true absences', i.e. they use a two-step approach where the selection of pseudo-absences is stratified according to an Habitat Suitability Index (measure of the potential habitat suitability for the species, computed from a profile technique such as ENFA).

Since the method of selection of the pseudo-absences strongly conditions the results obtained in the final model, the two-step approach is generally advised. Other crucial recommendations brought up to improve the model performance are: (1) eliminating buffered zones around presences, so that pseudo-absences are not drawn from (expected) suitable places for the species (Akçakaya and Atwood, 1997; Alexander et al., 2005; Olivier and Wotherspoon, 2006; Zarnetske et al., 2007; Ervin, 2009); (2) selecting an appropriate spatial subarea/extent or background size, i.e. neither too small nor too large, in order to avoid producing spurious results (VanDerWal et al., 2009); (3) minimizing the error experiment and using a 'sufficient' sample size (> 30 observations) for presence-only data sets (Guisan et al., 2007); and (4) sampling a number of pseudo-absences greater than the number of presences (Hengl et al., 2009; Warton and Shepherd, 2010).

5.2.2.2 Case-control studies

In a case-control study, cases events are those for which the outcome of interest has been observed (i.e. the disease, or the bicycle accident in our case) and controls are those in the same group/population (i.e. the cyclists) without the outcome of interest (Grimes and Schulz, 2005). Controls provide an estimation of the background frequency of an exposure in the study group, or population (i.e. the cyclists, or – ideally – the distance or travel time of trips carried out by these cyclists) (*ibid.*). As suggested in ecological modelling (Section 5.2.2.1), the use of an appropriate control group matters since a poor choice (of controls) can lead to wrong inference and, hence, to bad recommendations for policy makers and planners. According to Grimes and Schulz (2005), controls should be: (1) free of the outcome of the interest; (2) representative of the population at risk of the outcome, i.e. they should have the same risk of exposure as the cases; (3) selected independent of the exposure of interest. For a small number of cases, it is also suggested to draw up to four times controls in order to improve the power of the study. Beyond this ratio of 4/1, the improvements in the results (from the increase in the number of controls) are poor (*ibid.*).

Studies with case events only (i.e. without controls) often sample the controls from unknown or known population groups (e.g. the passengers of a cruise ship for who a disease of interest was not detected) (*ibid.*). In particular, in point process spatial models, researchers also commonly use the spatial distribution of another common outcome/disease as control group, which is assumed to reflect well the spatial distribution of the outcome/disease of interest (Diggle, 1990). For instance, Diggle (1990), Hossain and Lawson (2009) and Lawson (2009) used the cases of respiratory cancer of the lung as controls for modelling the spatial distribution of the cases of larynx cancer.

5.2.3 ... to traffic accident research

Literature in ecology and epidemiology provide well-founded methodological concepts that could be easily replicated to traffic accident research, for which only case events (i.e. road accidents) are registered. A case-control strategy is then opted here, accounting for some of the rigorous methods of selection of controls (or pseudo-absences) as implemented in group discrimination techniques. In particular, case events are here defined as being locations where a bicycle accident occurred on the network during a definite period of study (2006-2008), while controls are locations where no accident has been officially registered during the same period (2006-2008). Since the absence of accident is not absolutely certain for each control point (because of underreporting issues related to the registration of bicycle accidents), the concept of ‘pseudo-absences’ – as used in ecological modelling – could also be appropriate to refer to locations on the road network where no bicycle accident occurred. However, by convention, we decided to use the term ‘controls’ in this chapter.

5.2.3.1 Exposure variable

The only barrier to the replication of such a case-control strategy comes from the availability of an ***exposure variable***, from which controls can be selected/extracted as point events. Since the focus of this chapter is on bicycle accidents, the controls could be drawn from the places of residence of cyclists. Nevertheless, this is a somewhat naïve approach since the bicycle accidents are events that *result from the trips carried out by the cyclists*, and are hence expected to occur in greater *numbers* in places where the bicycle traffic is high (rather than in places where most of the cyclists live). As a result, an ideal exposure variable could be an estimation of the ***bicycle traffic*** (e.g. the total distance or time spent cycling, for each street of the road network). Unfortunately, except in some cohort studies or surveys, such a traffic variable is seldom available (Quddus, 2008), especially for non-motorised transport modes for which less attention is generally paid by planners, scientists or policy makers. Often, the best available data is the population of cyclists, aggregated by spatial units.

A solution proposed in this chapter in order to obtain an exposure variable is derived from the 'gravity-based' concepts as conceptualised in accessibility research (see e.g. Geertman and Ritsema van Eck, 1995; Geurs and Ritsema van Eck, 2001; Geurs and van Wee, 2004). In the literature, the 'gravity-based index' at location s, also called 'potential index' and noted P_s, is described by the following general form (Hansen, 1959; Geurs and Ritsema van Eck, 2001):

$$P_s = \sum_{t=1}^{T} a_t f(c_{st}) \tag{5.1}$$

where T is the number of spatial units (or locations), a_t are the 'opportunities' (e.g. number of activities, cyclists, etc.) in location t, c_{st} is the measure of spatial separation between s and t (e.g. the distance or travel time), and $f(c_{st})$ is the impedance function, denoting the deterrent effect of spatial separation between s and t ($s,t = 1, \ldots, T$). In other words, P_s is a measure of accessibility in s to all opportunities a in t, weighted by the spatial separation between s and t. Note that the impedance function can be of different forms (using e.g. power, exponential, Gaussian, logistic functions), and that its choice has a significant influence on the results (Geurs and Ritsema van Eck, 2001; Haynes et al., 2003). In particular, the negative exponential function $f(c_{st}) = exp(-\boldsymbol{\delta}.c_{st})$ (where δ is a non-negative parameter) is often preferred since it is the most closely tied function to the travel behaviour theory.

In this chapter, the potential index specification is adapted to estimate the potential bicycle traffic per spatial unit s, i.e. the (potential) background frequency of the exposure of cyclists to accidents. Such an adapted specification is here called the '***Potential Bicycle Traffic Index***' (noted PBTI; see Equation 5.15 for further details). Evidence in the literature lends strong support to the choice of such a potential index as proxy for the bicycle traffic, since it is often correlated to trip generation and closely reflects the actual behaviours in terms of the induced demand for travel (Haynes et al., 2003; Thill and Kim, 2005).

5.2.3.2 Selection of controls

Just as for ecological modelling, the random selection of controls is weighted/stratified as a function of the PBTI. Hence, the number of controls to be drawn varies from one spatial unit to another, proportionally to this index (i.e. in proportion to the bicycle traffic transiting in each statistical ward). In other words, the number of controls will be higher in areas where the (potential) bicycle traffic – i.e. the exposure – is higher (and inversely). Formally, the number of controls m_s to be drawn in spatial unit s is:

$$m_s = \frac{P_s^*}{\sum_{s=1}^{T} P_s^*} . M_0 = r_s . M_0 \tag{5.2}$$

where P_s^* is the adapted version of the potential index (i.e. the PBTI; see Equation 5.15), M_0 is the total number of controls, which is here four times greater than the number of (geocoded) accidents n_{acc} (as suggested by Grimes and Schulz (2005)). Since m_s is rounded to the closest integer value, it can be inferred that $M_0 \neq \sum_{s=1}^{T} m_s = M$ (where M is the total number of controls sampled in the studied area). Note finally that r_s is defined as the 'relative potential index' at location s and denotes the attractiveness at location s (or the relative potential intensity of the bicycle traffic at s) compared with all other locations.

Given that bicycle accidents generally happen on a road network, control points are constrained to be drawn on that network, at the exclusion of non-bikeable roads (e.g. tunnels) and linear buffered zones around the accidents in order to preclude the sampling from these zones. Such linear buffers correspond to black spots of accidents obtained from the network kernel density estimation method provided in SANET v.4 (Okabe et al., 2009). Further details on the definition of the 'bikeable network' as well as on the measure of the PBTI are provided in the section dedicated to the data collection (see Section 5.3.1.2).

5.2.4 Modelling strategy

The binary dependent variable used for modelling is derived from the combination of case events (occurrence of a bicycle accident at location i along the network) and controls (no bicycle accident at i). Case events are noted '1' and controls are noted '0'; this makes the use of logistic regression modelling possible if risk factors (dependent variables) are identified for both cases (1) and controls (0). Comparative statistics (i.e. chi-square tests, Wilcoxon tests, odds ratios, etc.) are first performed in order to examine if bicycle accidents are more likely (or not) to be associated with some specific risk factors. In a second step, logistic and intrinsic conditional autoregressive models are performed taking multicollinearity, heteroskedasticity and spatial autocorrelation into account. Modelling steps are conducted within a Bayesian framework and are described in the next subsections.

5.2.4.1 Bayes rule

Bayes rule (or Bayes' theorem) is the basis for Bayesian inference and can be simplified as follows when ignoring the normalising constant (Gelman et al., 1995; Kéry, 2010):

$$p(\theta|x) = \frac{p(\theta).p(x|\theta)}{p(x)} \propto p(\theta).p(x|\theta) \tag{5.3}$$

where θ is a vector of k parameters, x is a vector of n observations (i.e. the data), $p(\theta|x)$ is the posterior distribution of the parameters θ given the data x, $p(x|\theta)$ is the likelihood function of the data x given the parameters θ, and $p(\theta)$ is the prior distribution of the parameters θ (i.e. the prior beliefs). The posterior distribution is hence summarised as the product between the likelihood function and the prior distribution of the parameters. In other words, the analyst's understanding about the parameters θ is derived from combining the analyst's prior knowledge about the values of these parameters and the observed data (Wintle, 2003), with less emphasis placed on the prior knowledge if the observed data set is large (and inversely) (Gelman et al., 1995; Bolstad, 2007; LeSage and Pace, 2009).

5.2.4.2 Bayesian hierarchical modelling and accident risk model

Hierarchical Bayes allows to accommodate the inherent stochasticity of some models – such as this found in the spatial models – owing to its structure in several hierarchical stages (Congdon, 2003; Clark, 2005; Zhu et al., 2006; Bivand, 2008; Ntzoufras, 2009). The prior parameters are supposed to be random variables and may depend on distributions (the prior distributions) that may in turn depend on other parameters at a second level of the hierarchy. These latter parameters are called the 'hyperparameters', and may also have their own (hyperprior) distribution (Borgoni and Billari, 2003; Gelman et al., 1995; Ntzoufras, 2009).

Since the dependent variable is here binary, a two-stage conditional Bernoulli model with a logistic link is appropriate for predicting the probability of having a bicycle accident at location i (i = 1, ..., n; where $n = n_{acc}+M$):

$$y_i \sim Bernoulli(p_i) \tag{5.4}$$

$$logit(p_i) = \log\left[p_i/(1-p_i)\right] = \boldsymbol{\alpha} + \boldsymbol{x}_i\boldsymbol{\beta} \tag{5.5}$$

where y_i is the dependent variable ($y_i = 1$ if a bicycle accident occurred at location i; $y_i = 0$ otherwise), p_i is the probability of having a bicycle accident at location i, $\boldsymbol{\alpha}$ is the intercept of the model, $\boldsymbol{\beta}$ is the vector of parameters, and $\boldsymbol{x}_i$ is the vector of risk factors (explanatory variables). This is the first stage of the so-called '***accident risk model***' (noted 'Model 1' in the results). At this stage of the model, note that the risk factors $\boldsymbol{x}_i$ may be centered at zero in order to reduce correlations between the parameters. Interestingly, centering also allows increasing the speed of convergence and improves the inference of the model (Gelman and Hill, 2007; Best and Richardson, 2009).

At the second stage of the model, highly uninformative[3] prior distributions are generally assigned to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ when there is no prior information about the parameters. Such uninformative distributions reflect the prior ignorance we have about the parameters of the variable, and hence avoid strong prior beliefs about these latter so that the posterior distribution is unaffected by information external to the data (Gelman et al., 1995; Bolstad, 2007; Lawson, 2009). In general, (uninformative) normal distributions with mean 0 and precision 1.10^{-6} (precision = 1/variance) are specified for the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Formally, this is often noted in the literature as: $\boldsymbol{\alpha},\boldsymbol{\beta} \sim N(0,1.10^{-6})$, where N denotes the Normal/Gaussian prior distributions for the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, with mean μ = 0 and precision $\tau = 1.10^{-6}$.

[3] Also called flat, vague or diffuse prior distributions. Such uninformative prior distributions are used in order to 'let the data speak for themselves' (Gelman et al., 1995; Ntzoufras, 2009).

5.2.4.3 Autoregressive and autologistic risk models

Intrinsic Conditional Autoregressive (ICAR) model

Unobserved spatial effects – also termed ***random effects*** – can be incorporated in a statistical model to account for an extra quantity of variation (or unexplained variance) and then to avoid erroneous inferences regarding the parameter estimates[4] (Dormann et al., 2007; Miller et al., 2007). Two basic forms of random effects are generally distinguished: (1) the uncorrelated (or unstructured) heterogeneity, and (2) the correlated (or structured) heterogeneity. The uncorrelated heterogeneity refers to an independent and spatially uncorrelated form of extra variation (e.g. overdispersion), while the correlated heterogeneity implies that spatial autocorrelation exists between the spatial units (Lawson et al., 2003; Lawson, 2009). In particular, spatial autocorrelation could arise from the existence of unobserved effects (e.g. a key risk factor which is not included into the model) and/or from the fact that values at nearby locations depend from each other and hence are more (or less) similar/related than those further apart (e.g. spatial clustering of bicycle accidents) (Tobler, 1970; Dormann et al., 2007).

Incorporating random effects in the model provides a robust basis for inference when spatial autocorrelation and overdispersion are both present (see e.g. Borgoni and Billari, 2003; Miaou et al., 2003; Law and Haining, 2004; Zhu et al., 2006; Aguero-Valverde and Jovanis, 2006; Eksler, 2008; Eksler and Lassarre, 2008; Quddus, 2008; Haining et al., 2009; Lawson, 2009; Haque et al., 2010; Ishihama et al., 2010; Lord and Mannering, 2010). In Bayesian hierarchical modelling, an extension of the previous formulation (Equation 5.6) to the inclusion of such random effects is relatively straightforward given that all parameters are considered as being stochastic/random (Lawson et al., 2003; Bolstad, 2007; Lawson, 2009). Given that the bicycle accidents spatially concentrate on the Brussels' road network (see Chapter 4 and Section 5.3.1.2), a spatial Bayesian specification of the accident risk model (i.e. including the random effects) is here proposed. The first stage of such a model (Model 2) is formally defined as follows:

$$logit(p_i) = \boldsymbol{\alpha}' + \boldsymbol{x}_i\boldsymbol{\beta} + u_i + v_i \tag{5.6}$$

[4] For instance, if a statistical model using such (spatial) data still exhibits some spatial autocorrelation in its residuals, the assumption of independently and identically distributed (i.i.d.) residuals is violated, which may bias parameter estimates and result in increased type I errors rates (i.e. falsely rejecting the null hypothesis of no relationship between the dependent variable and the risk factors) (Dormann et al., 2007; Miller et al., 2007).

where u_i is the correlated heterogeneity and v_i is the uncorrelated heterogeneity associated with accident *i*. Overall, it is suggested to use both random effects since there is no prior knowledge about the form of the unobserved effects (Lawson, 2009). At the second stage of this spatial model, an improper[5] (flat) uniform prior distribution is assigned to α (Besag and Kooperberg, 1995; Thomas et al., 2004), while the parameters $\boldsymbol{\beta}$ are assumed to follow a normal distribution with mean 0 and precision 1.10^{-6} (i.e. $\boldsymbol{\beta} \sim N(0, \tau_\beta)$, where $\tau_\beta = 1.10^{-6}$). Focussing on the random effects, the prior distribution for the uncorrelated heterogeneity (v_i) is assumed to follow an uninformative normal $v_i \sim N(0, \tau_v)$, where τ_v is the precision of v_i. At a third stage of the model, the precision τ_v is assumed to follow a highly uninformative prior gamma distribution $Ga(0.5,0.0005)$ (Kelsall and Wakefield, 1999). Concerning the correlated heterogeneity, the spatial interactions between the neighbouring bicycle accidents are defined conditionally, with an Intrinsic Gaussian conditional autoregressive (ICAR) prior distribution being assigned for u_i (Besag et al., 1991):

$$[u_i \mid u_j,\ i \neq j] \sim N(\overline{u}_i, \tau_{u,i}) \tag{5.7}$$

where *j* is a neighbour of *i* (as defined in a binary spatial weight matrix), u_j is the correlated heterogeneity associated with accident *j*, and $\overline{u}_i$ as well as $\tau_{u,i}$ are defined as:

$$\overline{u}_i = \frac{1}{q_i}\sum_{j \neq i} u_j w_{ij} \tag{5.8}$$

$$\tau_{u,i} = \frac{\omega_u^2}{q_i} \tag{5.9}$$

where w_{ij} are the weights of the binary spatial weight matrix ($w_{ij} = 1$ if *i* and *j* are neighbours, $w_{ij} = 0$ otherwise), $q_i = \sum_j w_{ij}$ (which corresponds to the number of neighbours of accident *i*), and ω_u^2 is a parameter controlling the amount of variability in u_i. A prior gamma distribution $Ga(0.5,0.0005)$ is assigned to the inverse of ω_u^2 ($= 1/\omega_u^2$) at the third stage of this so-called **ICAR model** (Kelsall and Wakefield, 1999). Note finally that the spatial weight matrix is here symmetric and that two accidents *i* and *j* are 'neighbours'

[5] An improper distribution – in opposition to a proper distribution – refers to a distribution that does not integrate to 1. Note that a posterior distribution can be proper even when using an improper prior distribution (Lawson, 2009; Ntzoufras, 2009).

(i.e. $w_{ij} = 1$) if the network distance d_{ij} between these latter is lower than 100 meters. Otherwise, if $d_{ij} > 100$m, then $w_{ij} = 0$ (i.e. i and j are not neighbours)[6].

Autologistic model

Another spatial specification used in this chapter to predict the risk of bicycle accident is the **autologistic model** (Model 3). This model intends to capture the effect of spatial autocorrelation by including – at the first stage of the Bayesian hierarchy – an additional variable called the 'autocovariate' (Flahaut, 2004; Wintle and Bardos, 2006; Dormann, 2007; Dormann et al., 2007; Miller et al., 2007). Equation 5.5 is hence re-specified as follows (Besag, 1974):

$$logit(p_i) = \boldsymbol{\alpha}' + \boldsymbol{x_i\beta} + \lambda S_i \quad (5.10)$$

where S_i is the autocovariate for the bicycle accident i and λ is the parameter for the autocovariate. Such an autocovariate is generally defined as a weighted sum (or average) of the observations in the neighbourhood (Wintle and Bardos, 2006; Dormann et al., 2007):

$$S_i = \sum_{j \neq i} w_{ij}^* y_j^* \text{ (weighted sum), or } S_i = \frac{\sum_{j \neq i} w_{ij}^* y_j^*}{\sum_{j \neq i} w_{ij}^*} \text{ (weighted average)} \quad (5.11)$$

where w_{ij}^* are the weights assumed to represent the relationship existing between i and its neighbours j, and y_j^* are the response values observed for these neighbours (j). Note that the spatial weight matrix has here a more complex and flexible definition compared to this used in the ICAR model. In order to optimise the model inference, several specifications were indeed tested – through trial and error – using different functional forms for the network distance between i and j. It then turned out that a spatial weight matrix accounting for a distance-based relationship between the bicycle accident i and the nearest bicycle accidents j (1st order neighbourhood) was the best to capture the unexplained variance associated with the presence of spatial autocorrelation: $w_{ij}^* = e^{-d_{ij}}$ if j is a 1st order neighbour/accident, $w_{ij}^* = 0$ otherwise (j is a 2nd order neighbour/accident or more).

Similarly to the logistic specification, highly uninformative (normal) prior distributions are here selected at the second level of the hierarchy ($\boldsymbol{\alpha},\boldsymbol{\beta},\lambda$ ~

[6] Note that another spatial weight matrix (1st order contiguity) – based on a Network Voronoi diagram computed in SANET v.4 (Okabe et al., 2009) – was also tested, but did not improve the results.

$N(0,1.10^{-6})$). Although autologistic models use a likelihood approximation to the maximum likelihood method (i.e. a pseudolikelihood approximation), they generally provide a better estimation than the basic logistic regression (Augustin et al., 1996; Hoeting et al., 2000) as well as a reasonable approximation when spatial autocorrelation is relatively low (Lawson, 2009). Moreover, the Bayesian inference may still be improved by extending the autologistic model to include random effects (e.g. in the form of uncorrelated heterogeneity) (*ibid.*).

5.2.4.4 Initial values and model selection

As illustrated in Figure 5.1, model selection is carried out using a three-step approach, aiming at (1) evaluating the statistical fit of a wide range of multivariate (auto-)logistic models (and diagnosing these for the presence of statistical biases) within a frequentist framework, (2) selecting the risk factors and initial values of the Bayesian models on the basis of the frequentist inference, and (3) evaluating the statistical fit of the Bayesian models (and diagnosing these for convergence).

In the first step, logistic and autologistic regressions are performed and evaluated within a frequentist framework in order to get the initial values. An overall model evaluation of these frequentist models is carried out using: (1) inferential statistical tests (Likelihood ratio and Wald test) in order to analyse the significance of the model, compared with the intercept-only/null model; (2) statistical tests of the individual parameters of the risk factors (Wald chi-square statistic); (3) goodness-of-fit statistics (i.e. Log Likelihood (LL), Akaike's Information Criterion (AIC)) and tests (i.e. Hosmer-Lemeshow test (HL), and Le Cessie-Houwelingen test (LCH)); and (4) validations of predicted probabilities, using the *c* statistic and misclassification rates (cut-off value: 0.5) (see e.g. Joanne-Peng et al. (2002) for further details on some of these statistics). Also note that diagnostics for multicollinearity (i.e. Variance Inflation Factors (VIF), Condition indices (CI)), spatial dependence of the dependent variable (join-count test statistics under non-free sampling, with or without adjustment for no-neighbour observations) and spatial autocorrelation of the residuals (Moran's *I*) [7] are also performed and influenced our choice of the risk factors in the models. Finally, heteroskedasticity – when present – is corrected by implementing the Huber-White method.

[7] Note however that Moran's *I* index is generally not recommended for logistic regression modelling, since its statistical basis for inference is still not well-founded. In this case, the analysis of Moran's *I* – and the attendant conclusions of the test about an eventual detection of spatial autocorrelation – is then prone to particular caution.

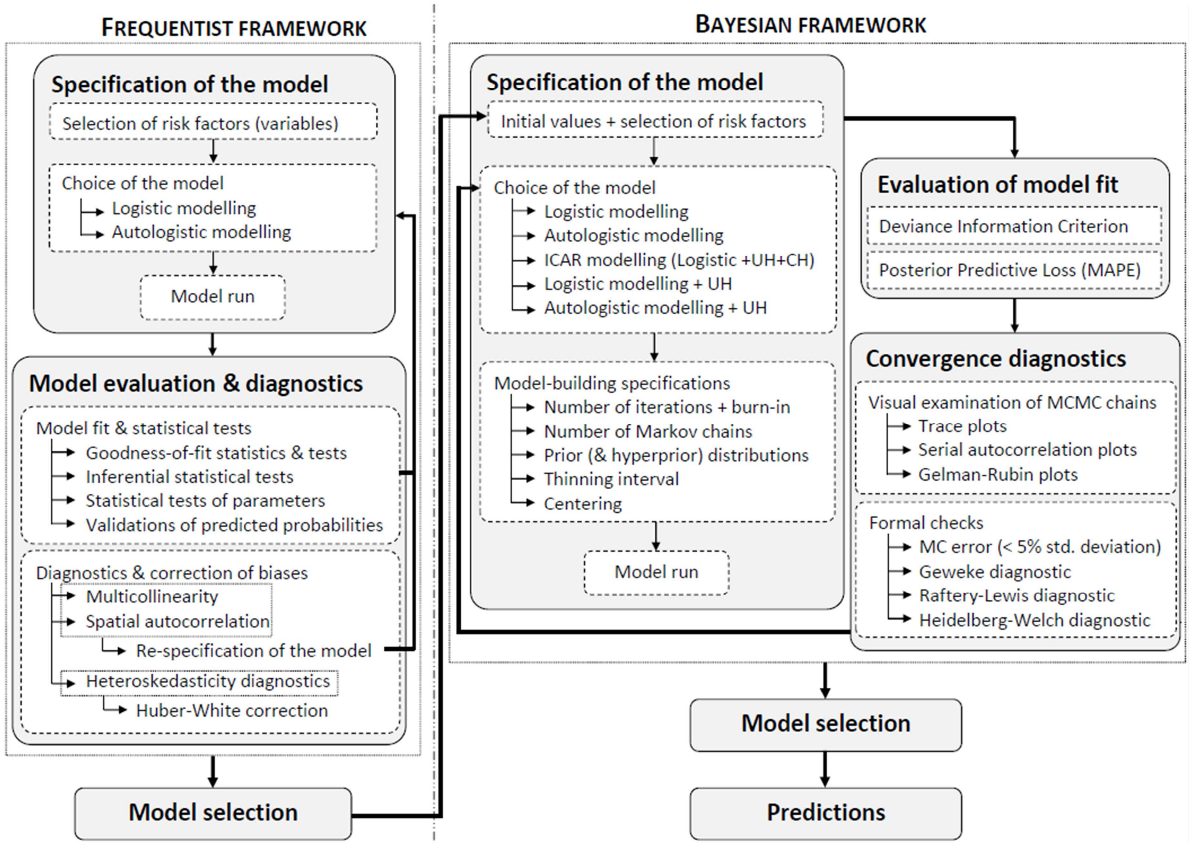


Figure 5.1: Modelling strategy (UH: uncorrelated heterogeneity; CH: correlated heterogeneity)

In the second step, the initial values of the Bayesian models are determined on the basis of the parameter estimates we obtained for the 'best' frequentist models. The Bayesian models were used instead of the frequentist ones because they outperformed these latter in terms of inference; the frequentist models were then only used to get some prior insight/knowledge on what could be appropriate initial values for the parameters of the risk factors (estimated within a Bayesian framework). Moreover, starting from initial values that are close to the actual (or 'true') values of the parameters has also the advantage to speed up the convergence of the (Bayesian) models (or, at least, avoid an 'accident' in the model because of e.g. numerical difficulties in sampling).

In the last step, the statistical fit of the Bayesian models is computed in order to compare their performance and select a better-fitting model. Two goodness-of-fit measures are applied to compare different models: (1) the Deviance Information Criterion (DIC), and (2) the Mean Absolute Predictive Error (MAPE) (Lawson, 2009). DIC is often suggested as an adapted measure to compare the fit and complexity of the hierarchical Bayesian models, for which the exact (or effective) number of parameters is not always clearly defined (Spiegelhalter et al., 2002; Law and Haining, 2004; Lawson, 2009; Ntzoufras, 2009). The DIC is a generalisation of the AIC and is defined as follows:

$$DIC = D(\overline{\theta}) + 2p_D = \overline{D} + p_D \tag{5.12}$$

where $D(\overline{\theta})$ is the deviance evaluated at the posterior means of the parameters ($\overline{\theta}$), p_D is the effective number of parameters in the model, and $\overline{D}$ is the posterior mean of the deviance (for Bernoulli likelihood, the deviance is: $-2\sum_i [y_i \log(p_i) + (1-y_i)\log(1-p_i)]$ (Law and Haining, 2004)). DIC hence expresses a trade-off between the model fit (measured by $D(\overline{\theta})$) and the model complexity (measured by p_D, which acts as a penalty for the model fit as the number of parameters increases) (Law et al., 2006; Lawson, 2009; Kéry, 2010). As with AIC or BIC, models with lower DIC values indicate better fitting models and are hence preferred.

MAPE is a posterior predictive loss (PPL) measure, aiming at comparing the predictive ability of the models (while the DIC estimates how well the model fits the observations). It is particularly useful for binary data as it computes the proportionate misclassification under the fitted models (Gelfand and Gosh, 1998; Lawson, 2009). It is defined as follows:

$$MAPE = \sum_q \sum_l \left| y_q - \hat{y}_{ql} \right| \Big/ (G \times n) \tag{5.13}$$

where y_q is the qth observed value, $\hat{y}_{ql}$ is the qth predicted value given the parameters at the lth iteration, G is the sampler sample size and n is the number of observations. Models with smaller values of MAPE are preferred.

5.2.4.5 Convergence diagnostics (Bayesian framework)

The Markov chain starts from an initial value attributed by the analyst to each parameter, which is either arbitrary or computed from a frequentist method (in order to avoid a long computation time). After a suitable number of iterations (ensuring that the chain is independent of the initial values), the chain is expected to reach an equilibrium distribution, or convergence. The first draws obtained before convergence are called the 'burn-in period' and are discarded since they are not representative of the equilibrium distribution (Gelman and Hill, 2007; Bivand, 2008; Lawson, 2009; Ntzoufras, 2009; Kéry, 2010). Summary statistics of the posterior distribution are then computed directly from the remaining simulations.

In order to provide evidence for the robustness of convergence, most model estimations should run at least two or three chains in parallel (having different initial values) as well as a 'sufficient' number of iterations for monitoring convergence. Nevertheless, it has the disadvantage to require more computation time (Law et al., 2006; Gelman and Hill, 2007; Lawson, 2009). Given that the burn-in period can vary considerably from one MCMC estimation to another, the convergence of the chain is not guaranteed and must be monitored performing a qualitative judgement using several convergence diagnostics (Lawson, 2009). In a first step, such diagnostics may consist of a simple visual examination of the MCMC chains (also called 'trace plots of the samples'), serial autocorrelation plots, and the Gelman-Rubin statistic plots. MCMC chains should look like oscillograms stabilising around a mean value without any tendency or periodicity, hence indicating that a good mixing is obtained (Law et al., 2006; Ntzoufras, 2009; Kéry, 2010). The overall autocorrelation between successive values (sampled from the posterior distributions of the parameters) may also be examined using the autocorrelation plots. In particular, a high autocorrelation value (i.e. near to 1) indicates that the samples are dependent, or − in other words − is suggestive of a slow mixing of the chain. Such a problem is overcome by 'thinning' the chain (i.e. considering a sampling lag) and keeping only the first every k iterations (where k is the thinning parameter). This strategy has also the advantage to reduce the computing time and save storage space (Wintle, 2003; Sturtz et al., 2005; Ntzoufras, 2009). Finally, convergence can be assessed analysing the Gelman-Rubin statistic plots. Implemented when multiple chains are run in parallel (starting from different initial values), this test statistic compares the between- and within-chain variance (like an ANOVA-type diagnostic test). Convergence is likely to be achieved when values close to 1 and lower than 1.1 are obtained (Spiegelhalter et al., 2003; Lawson, 2009; Ntzoufras, 2009; Kéry, 2010).

In a second step, more formal checks are required to monitor convergence (Geweke, 1992; Gelman and Rubin, 1992; Raftery and Lewis, 1992; Heidelberg and Welch, 1992). First, the Monte Carlo (MC) error of the posterior mean must be assessed for each parameter. MC error measures the variability of each parameter due to the simulation, i.e. the sampling error in this simulation (Ntzoufras, 2009; Kéry, 2010). As a rule of thumb, this statistic should be <5% of the posterior standard deviation of a parameter, which is generally the case for large and independent samples (Law and Haining, 2004; Law et al., 2006; Kéry, 2010). Second, the Geweke statistic is computed as a score test based on the comparison of the means of the beginning and the end of a single chain. A test statistic (Z) with values $|Z| < 1.96$ suggests that both means are equal, and hence that the chain has converged (Geweke, 1992; Smith, 2005; Bivand, 2008; Lawson, 2009; Ntzoufras, 2009). A third test is the Raftery-Lewis diagnostic, which is also applied on a single chain of a parameter to evaluate the required thinning interval as well as the minimum number of iterations to achieve a pre-specified level of accuracy. The required thinning interval is roughly estimated by the dependence factor (I), for which values greater than 5 indicate convergence failure and suggest the need to reparameterise the model (Geweke, 1992; Raftery and Lewis, 1992; Smith, 2005; Ntzoufras, 2009). Last but not least, the Heidelberg-Welch diagnostic (Heidelberg and Welch, 1992), which is also used for the analysis of single chains, consists of a two-part test (stationarity and halfwidth tests). In the first part of the test, the stationarity of the chain is monitored using the values from an MCMC output. Without evidence of stationarity, the test is repeated on a reduced sample (the first 10% of the iterations are dropped) until the resulting chain passes the test or more than 50% of the iterations are discarded (Smith, 2005; Ntzoufras, 2009). If the test is rejected, a longer run is required to achieve convergence. Otherwise, the halfwidth test (second part of the diagnostic) is run on the portion of the chain that passed the stationarity test. If the halfwidth test is passed, the required precision of the posterior mean (of the parameter of interest) is achieved. In the opposite case (failure of the test and low accuracy of the mean), a longer run of the chain must be considered to reach the required precision (Smith, 2005; Ntzoufras, 2009).

Finally, the last step consists of: (1) analysing the Bayesian residuals and diagnosing if spatial autocorrelation is still present (using the Moran's I index); (2) summarizing the posterior distributions of the parameters, using point estimates of these latter (e.g. posterior mean and 95% credible intervals of the parameters); and (3) computing predictions for 'unobserved' locations and data (i.e. where accidents were not reported) in order to examine what could have been the risk of having a bicycle accident at such places during the period of study (2006-2008).

5.3 Data collection

Figure 5.2 illustrates the data collection step. First, an exhaustive review of the literature was conducted as regards risk factors that are likely to be (in-)directly associated with the occurrence of the bicycle accidents. Second, data were collected based on this review and on the knowledge provided by the 'grey' literature (e.g. reports based on the experience of cyclists or on findings of road safety institutes). Such data were either pre-processed from tabular data (or from other formats, e.g. pdf) with the aim to obtain spatial/geographic data, or digitized through a time-consuming manual process in a Geographic Information System (GIS) from orthophotos and maps. While digitizing the data, special attention was paid to the direction, year and type of spatial data (e.g. cycle facilities), thus allowing a categorization of these latter. As regards the dependent variable y_i, the bicycle accidents (cases) were geocoded along the road network and these data were completed using a set of controls, i.e. locations where a bicycle accident is not expected to have been occurred. Controls were generated using a stratified random sampling from an exposure variable (i.e. the PBTI). Third, the risk factors result from crossing the digitized spatial data with the binary dependent variable y_i into a GIS, and from manually checking the results of these crossings (only for $y_i = 1$). Fourth, the final database with accidents, controls and their respective risk factors is used for modelling the risk of having a bicycle accident along the Brussels' road network. By trials and errors (using diagnostic and goodness-of-fit statistics), the best accident risk models were selected and then used in order to compute predictions for a specific 'bikeable' trajectory of the network.

Most of the spatial data related to the road network and to the risk factors are provided by the Brussels Regional Informatics Center (BRIC), using the Brussels UrbIS[1] database. As mentioned in Chapter 4, 'unbikeable' links are excluded from the network data set in order to construct a 'bikeable' road network. This latter is modelled as a connected 'non-planar' graph, i.e. in such a way that the

[1] UrbIS is the acronym for 'Urban Information System'.

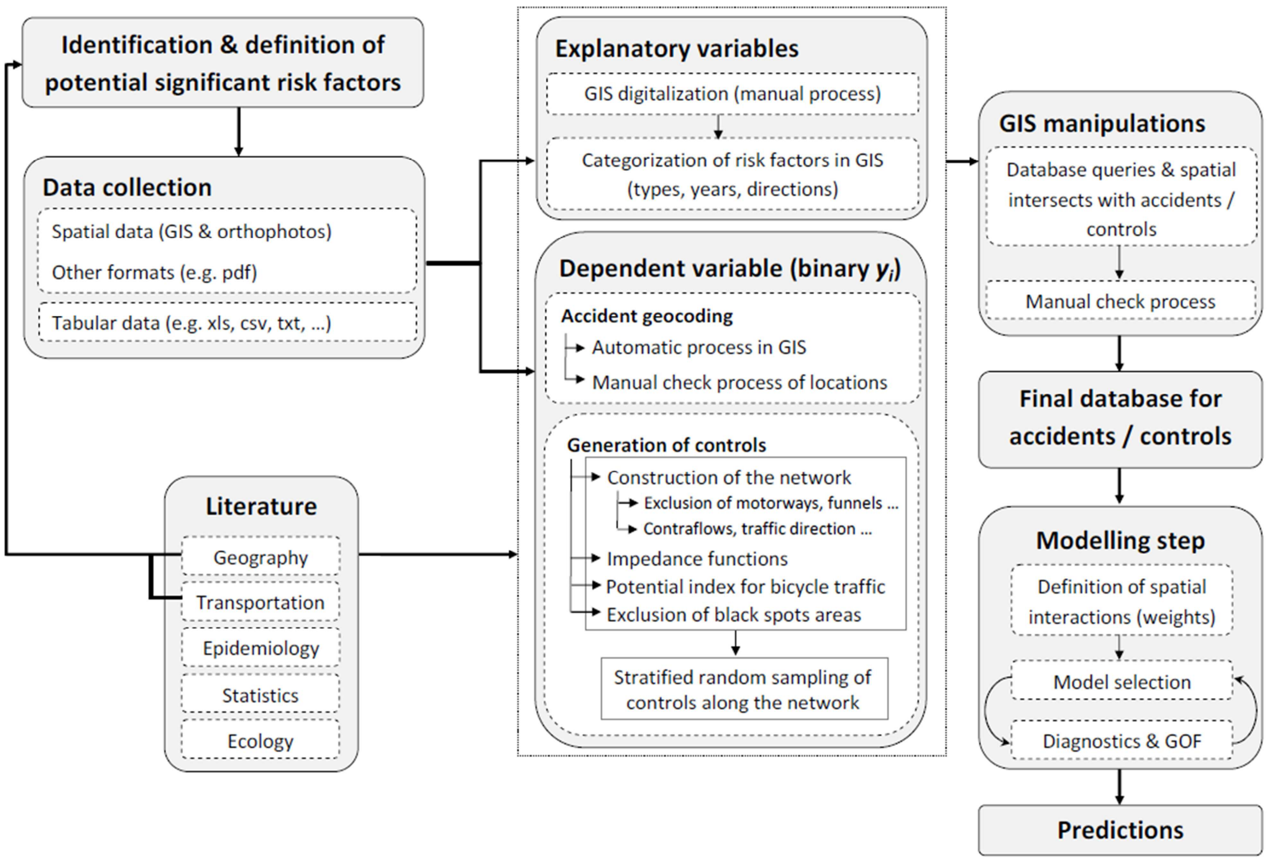


Figure 5.2: Data collection – conceptual framework

relative heights of road links (elevations) are considered when bridges and tunnels are present. It is also 'directed' in the sense that directions of travel along the road links are taken into account when computing the network distances. In particular, one-way streets and contraflows cycling are manually and visually identified owing to orthophotos (BRIC, Google Earth), cycling maps for the period 2006-2008 (Brussels Mobility) and data provided by BRIC.

In order to account for cyclists living in the Brussels' periphery (and hence avoid edge effects in the estimation of the PBTI), the road network is 35 km buffered around the BCR boundaries. This buffered network is defined in the same way as that of the BCR, but it does not include any information on contraflows cycling (not available). It is however not expected to affect the distance estimations significantly[1].

5.3.1 Accident data

5.3.1.1 Accident geocoding ($y_i = 1$)

Road accidents collected from the Directorate-General Statistics and Economic Information (DGSEI) are used within the framework of this chapter. A total of 644 bicycle accidents was censused in the BCR over the period 2006-2008. About 93% of these (600) are successfully geocoded using a semi-automatic process (see Chapter 4). Note that the severity of the accident is not considered here, which is however not a major limitation since most cycling accidents (95%) resulted in slight injuries. Regardless of the official statistics, the proportion of cycling accidents with slight injuries in the total number of cycling accidents is even expected to be higher. As mentioned in Chapter 4, accidents resulting in slight injuries (and/or with material damages) are indeed strongly underreported compared to the other degrees of severity (it is estimated that about 15% of the cycling accidents are reported by official statistics).

5.3.1.2 PBTI and generation of controls ($y_i = 0$)

As mentioned in Section 5.2.3.1, controls are randomly selected along the 'bikeable' network, but their sampling is stratified per spatial unit s (statistical wards) as a function of the potential bicycle traffic (here: the PBTI). Data from the 2001 Socio-Economic Census (DGSEI) are used to estimate the PBTI for

[1] As an illustration, the relative difference between network distances computed without any information on contraflows cycling and network distances with such information is on average low (1%) for the BCR.

Brussels. This latter is here noted P_s^* and uses a modified negative exponential function as impedance[2]:

$$P_s^* = a_s + \sum_{t=1}^{T} a_t . \left[\zeta_t . \exp(-\delta_t . d_{st}) + \eta_t . d_{st} . \exp(-\varepsilon_t . d_{st}) \right] \qquad (5.14)$$

where t are the statistical wards in the neighbourhood of s (s,t = 1,..., T and $t \neq s$), a_s (or a_t) is the number of cyclists commuting to work or school and living in s (or t), d_{st} is the distance measured along the 'bikeable' network (expressed in kilometres) between the centroids of s and t, and ζ_t, δ_t, η_t, ε_t are parameters of the impedance function attributed to the statistical ward t. Note that we here limit to commuter cyclists given that: (1) data on cycling trips carried out for other purposes (e.g. leisure, shopping) are not available; (2) considering commuting trips only is more robust in the sense it excludes the high variation that could be associated with occasional cycling trips (e.g. recreational trips). Only regular cyclists are hence taken into account to compute the PBTI. Also note that Equation 5.15 assumes that all statistical wards t have a same level of attraction, and thus that there is no preferential direction of travel for cyclists. This could be not the case in reality, with the town centre being the most attractive place in general. Nevertheless, many activities are located outside the town centre (leisure areas, corner shops, cultural centres, workplaces in industrial or business parks, etc.) and attract cycling trips, thus diverting these from the town centre.

For each (former) municipality o (each containing several statistical wards t), the parameters of the impedance function (ζ_t, δ_t, η_t, ε_t) are empirically calibrated on the basis of an observed impedance function of cycling trips (i.e. the observed proportion of cycling trips as a function of the distance). In other words, an impedance function is constructed for *each* municipality o, and then the values of the parameters calibrated at this scale are assigned to all statistical wards t that are contained in o. Such a method hence assumes that travel behaviours are different according to the place of residence of the cyclists. As suggested by exploratory analyses (not shown here), cyclists living in the town centre (CBD) indeed travel shorter distances than those living in more peripheral locations of the urban area (the availability and the proximity of facilities is higher in the town centre, which reduces the cycling distances of the commuter cyclists living there). Note that the parameters of the observed impedance functions are computed at the scale of the municipalities because the number of observations

[2] Such modified negative exponential functions turned out to provide a better fit to observed values, compared with a wide range of polynomial functions. The negative exponential function is here selected as it is the most tied function to the individuals' travel behaviour (Geurs and Ritsema van Eck, 2001).

(and hence the statistical significance) is higher at this scale than when working with statistical wards. In case that there are less than 30 observations for one municipality, interpolation of the number of cyclists is performed on the neighbouring municipalities (1st order queen contiguity). Finally, a 30 km 'guard area' (i.e. a buffer area that is external to the area of interest) was also used in order to avoid eventual 'edge effects' by accounting for the commuters that live in peripheral municipalities (in Flanders and Wallonia) and that could be likely to cycle into Brussels from these. The implementation of such a 'guard area' also prevents from considering Brussels as a 'closed system', since interactions exist within its urban region (which extends outside the administrative boundaries). As a result, about 550 functional forms of impedance functions (i.e. one for each municipality of the BCR and its guard area) were calibrated on the basis of observed travel behaviours of cyclists, thus leading to a better fit than when using only one functional form (Iacono et al., 2010).

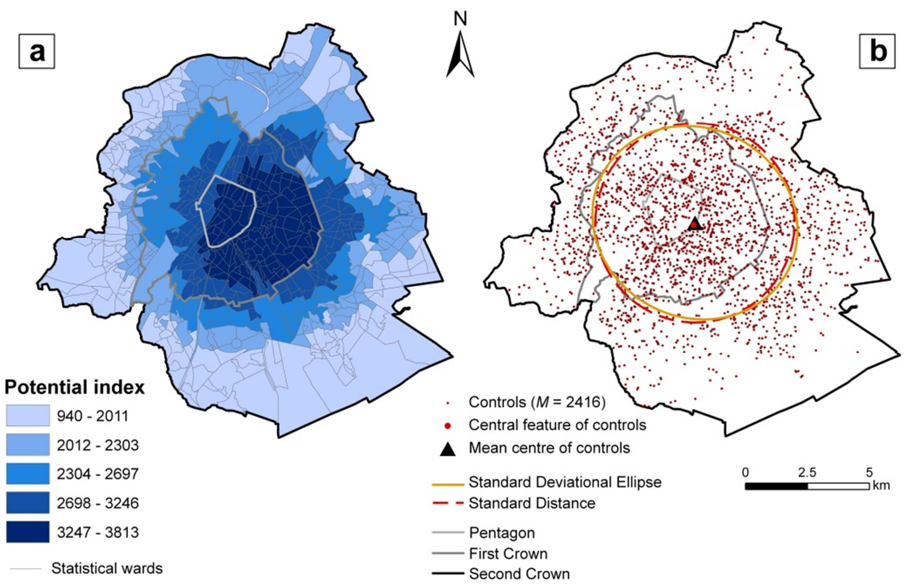


Figure 5.3: Spatial distribution of: (a) the exposure variable, i.e. the Potential Bicycle Traffic Index (PBTI), (b) control points, generated from the PBTI and drawn along the 'bikeable' network (without black zones). Data source: DGSEI.

Once the impedance functions calibrated, the PBTI is computed and provides an estimation of the potential number of cyclists stopping or transiting in *s*. Concretely, this hence refers to the number of cyclists living in *s* plus a number of cyclists living in the neighbourhood and being likely to travel the distance between their place of residence *t* and *s*. A visual check of Figure 5.3a suggests that the PBTI is close to the actual spatial patterns of the bicycle traffic, despite the fact that no preferential direction is assumed for cycling trips in Equation 5.15. Interestingly, the locations where large numbers of cyclists are reported by the yearly bicycle traffic counts (e.g. European district) all correspond to maximum values of the PBTI. Figure 5.3a also exhibits high PBTI values for the eastern parts of the so-called 'Pentagon' and 'First Crown' areas (areas delineated by major roads), which corresponds to the places where cycling trips and accidents are the most common in Brussels (see Chapter 4). At the opposite, low values are obtained for the southern and western parts of the BCR, which is an expected result since few cyclists are observed in these areas. As a last check, measures of central tendency (e.g. mean center, central feature) and spatial dispersion (e.g. standard distance, standard deviational ellipse) seem to confirm the validity of the results (Figure 5.3b) (see Levine et al. (1995) & Myint (2008) for further details about these measures).

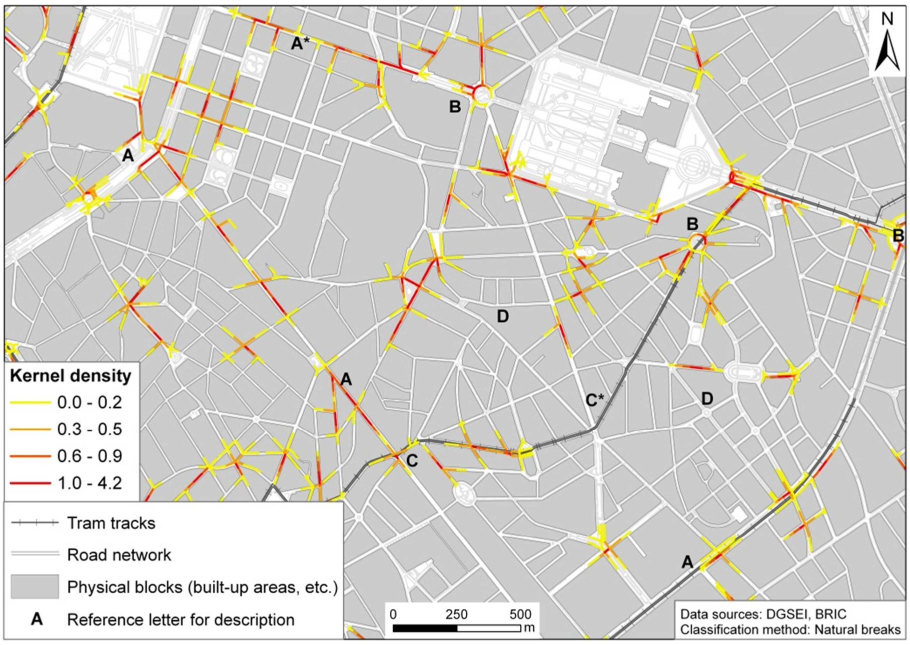


Figure 5.4: Black spots of bicycle accidents (2006-2008) in the Brussels' European district. A: major roads, with high capacity and dense motorised traffic volume; A*: *idem*, but with a separated cycle facility; B: large roundabouts, with dense motorised traffic volume; C: road with tram tracks (here: on-road and crossable); D: residential wards, with traffic-calming measures (speed humps, 30 km/h areas, etc.).

The number of controls to be drawn in each statistical ward s is now weighted as a function of the PBTI (see Equation 5.2). The higher the PBTI, the higher the number of controls to be drawn in s. Given that $M_0 = 4.n_{acc} = 2400$ (where n_{acc} is the number of bicycle accidents in Brussels), the total number of controls m_s is 2416. These latter are then drawn at locations i along the bikeable road network (stratified per s; i is contained in s), from which we removed the black spots of bicycle accidents (which is somewhat equivalent to the buffered areas in ecological modelling) in order to preclude the sampling from the close vicinity of bicycle accidents. Such black spots are obtained by performing a Network Kernel Density Estimation provided by SANET v.4 (Okabe et al., 2009), or simply by computing linear/network buffers (100m) around the bicycle accidents (e.g. using the Network Analyst extension of ArcGIS 9.x). In the former case, just note that a 100m bandwidth is used and that the presence of bridges and tunnels is taken into account (through a manual correction). Figure 5.4 provides an illustration of black spots for cyclists in the European district. Unsurprisingly, cycling accidents are more likely to be observed at intersections (especially at roundabouts) and on major roads with dense motorised traffic and/or tram tracks. At the opposite, it seems that residential roads are less prone to generate cycling accidents (which could be explained either by the traffic-calming measures, or by a greater degree of underreporting of cycling accidents in these streets (as suggested in Chapter 4)).

As a final step, a year (2006, 2007 or 2008) and a traffic direction are randomly assigned to the controls. This allows to associate the controls with the spatial risk factors, since these latter may be built at a definite moment over the period of interest (e.g. in 2008) or may be reported at one street side only (depending on the location where the control is located). Controls (noted $y_i = 0$) are finally appended to the geocoded bicycle accidents ($y_i = 1$) in the same database, which then makes possible the use of logistic regression modelling.

5.3.2 Risk factors

Classically, road accidents in general result from the interaction and combination between five categories of risk factors: human factors (e.g. driver behaviour, driver error), vehicle-related factors (e.g. size or state of the vehicle), infrastructure factors (e.g. crossroad design, pavement type), traffic conditions (e.g. density, speed), and environmental factors (e.g. lighting, weather) (Miaou et al., 2003; Li et al., 2007; BRSI, 2008). In this chapter, we mainly focus on infrastructure factors, traffic conditions and – to a lesser extent – environmental conditions. The other factors are however not available for the controls (as well as for the bicycle accidents (e.g. driver errors)). Appendix D.1 lists all risk factors used in this chapter as well as their definition, units and data sources. All of these data are collected for the period 2006-2008 and at the scale of the Brussels-Capital Region. Note that the list is quite exhaustive (> 45 risk factors) because we aimed at monitoring if some (potential) risk factors might have had an unexpected relationship with the locations where bicycle accidents occur.

5.3.2.1 Infrastructure factors

A first data set of infrastructure factors is collected within the framework of Chapter 4 and is here also exploited for statistical modelling. Contrary to this previous work, the presence, evolution and street side where the infrastructures are built are controlled over the 2006-2008 period, using orthophotos (BRIC, Google Earth) and accident data (DGSEI). This first data set collects data on: bridges/tunnels, traffic-calming areas, intersections, tram tracks and public transport stops, cycle facilities and discontinuities along the bicycle network, parking facilities (for motorised vehicles), proximity to activities and public services, and streets where contraflow cycling is permitted (see Chapter 4 for further information). Some modifications (i.e. changes in the definition of the factors, collection of additional variables, etc.) are however implemented here and described below, before outlining some additional risk factors created for the purpose of this chapter.

Updates

(1) Intersections are modelled as '**zones**' instead of points, as the probability a control falls exactly on a point (here: the intersection) is almost null. This also allows reflecting the zone/extent of the intersection in a more realistic way. For instance, in cases where the control is drawn 1m from the exact intersection point, it is indeed more realistic to consider that such a control belongs to this intersection (or 'intersection zone'). As regards roundabouts and traffic lights, the intersections are manually delineated on the basis of specific road features identified in the crossroad (e.g. stop lines, yield lines for roundabouts, etc.). Concerning the other types of intersections, the zones are defined as being 10m-length[3] linear buffers starting outwards in all (possible) network directions from the exact intersection point.

(2) Two other proxies are used with regard to intersections: the proximity to the closest intersection and the complexity index. First, network distances

[3] By convention, a 20m length is generally chosen (Liu and Jarrett, 2008), but this is expected to be not appropriate here since we manually digitized the roundabouts and signalized intersections (which have larger lengths) separately.

are computed between each accident / control and the **closest intersection** on the road network (whatever the type of intersection). Such a measure is particularly useful in the case where intersections would have a gradual (decreasing) influence with increasing distances, rather than a constant influence over a specified distance (e.g. 10m). Second, a **complexity index** is computed for each accident / control. Such an index consists of the sum of all road links starting / radiating outwards (in all possible network directions) from the accident / control location, over a certain distance or 'bandwidth'. Here the bandwidth values are 10, 20, 30, 40, 50, 75 and 100m. For instance, if the bandwidth is 10m, the complexity index computed for an accident located in a four-legs intersection is 40 m, whereas it will be 20m when occurring in the middle of a road link (if this latter is at least 20m length). Note that complexity is here defined in line with the Elvik's law on complexity (i.e. in the sense that it is a proxy for road legibility) (Elvik, 2006).

(3) One additional factor is collected with regard to the (public) **transport infrastructures**. It simply measures the network distance between the accident / control and the closest public transport stop. Concerning tram tracks, it is also not mentioned here if the cyclists actually crossed the tracks (e.g. at right angles, etc.) or if they cycled parallel to these. Indeed, the accident reports often give insufficient and/or imprecise information to infer what might have been the cyclist's direction relatively to tram tracks.

(4) Regarding the cycle facilities and the discontinuities along the network, two additional variables are created. The first one is created on the basis of a manual selection of the **cycle facilities located nearby parking areas**, i.e. at a distance < 0.9 m from the boundaries of the closest longitudinal parking area (which corresponds to the door zone) or public parking exit (i.e. other than a garage). The second consists in the network distance computed between each accident/control and the **closest discontinuity** (i.e. end or cut of a cycle facility along the bicycle network).

(5) As regards **parking facilities** (**function-based**), a specific type observed in some parts of the street should not be generalized to the whole street, especially if curb extensions reduce the space dedicated to parking (e.g. at intersections or at pedestrian crossings level). It would indeed be erroneously inferred that parking facilities would have had a role in the occurrence of the accidents reported in front of such curb extensions. A time-consuming digitization process is hence performed to account for all possible discontinuities (e.g. curb extensions) in the parking facilities observed along the road network. Network distances are then computed between each accident/control and the closest parking facility of each type,

hence providing a measure of proximity to these latter in order to explore their respective safety effect.

(6) Information on traffic direction in streets where **contraflow cycling** is allowed (or prohibited) is encoded in order to precise if the cyclist was facing motorised traffic or not.

(7) As regards **activities and public services**, network distances are computed from each accident/control to each of these activities/services in order to explore whether the proximity to these latter has an impact on the risk of having a bicycle accident. Several 'proximity variables' are hence obtained (i.e. one for each kind of service or activity).

Garage entrances/exits

Garage entrances and exits differ from public parking facilities (aspect- and function-based; see Chapter 4) in the sense that they mainly consist in *private* driveways and entrances to parking shops, shopping centres, firms or companies. They are expected to increase risk, especially when the driver leaves the garage/parking driveways to join a road or crosses a cycle facility while leaving/going into these. The presence of parked cars or other hurdles (e.g. plant tubs) may indeed hinder the field of vision of motorists and cyclists, which hence increases the risk of accident for both road users. In the literature, there is little evidence about a potential effect of garage entrances/exits on the occurrence of bicycle accidents. Rifaat et al. (2011) concluded that cyclists experienced higher non-injury and fatal risks on private driveways and parking lots. On the one hand, non-injuries result from collisions with hurdles and from the low speed of vehicles leaving/going into private driveways, while – on the other hand – fatal injuries are the effect of collisions of motorised vehicles with small children cycling/playing in the street (and being more vulnerable to road accidents). Greibe (2003) also found an inverted 'U-shape' relation between the number of accesses (private driveways, parking facilities, etc.) and accident risk, whatever the road user. Lower accident risks were found for roads with no access or with a large number of accesses, while the highest risks were reported for roads with a moderate number of accesses.

As regards to data collection, centroids of garage entrance/exits (modelled as lines into GIS) and are subsequently used to compute 3 indices measuring how 'frequent' the garage entrances/exits are in the close neighbourhood of bicycle accidents/controls. These indices are: (1) the number of garage entrances/exits within 100m (network distance) from the place of the accident/control (for the sake of brevity, note that ranges are here defined: 0, 1-10, 11-20…, and more than 70 entrances-exits); (2) the presence of at least one garage within 10, 50 or 100m (network distance) from the accident/control; (3) the network distance (in meters) between each accident/control and the closest garage entrance/exit. Finally, a fourth index is directly computed from the Brussels UrbIS database (BRIC) and represents the total garage length (i.e. the sum of all individual garage lengths) summed over a 100m network-based distance from the place of the accident/control. Let us note here that the incoming/outgoing flows of motorised vehicles are not available, which constitutes a limitation as the accident risk is likely to be higher for entrances to parking of shops and supermarkets (compared to private driveways of dwellings).

Proximity to the town centre

The number and risk of bicycle accidents are expected to increase while coming closer to the town centre. This latter is indeed very dense in population, jobs and activities (e.g. shops, work, schools, etc.) and hence is more prone to generate a high number of conflicting situations and accidents since it attracts high volumes of motorized and non-motorized traffic compared to other parts of the urban area. Irrespective of the type of road user, Levine et al. (1995) indeed found that most accidents occurred in the employment areas, while Greibe (2003) observed that accident risks were the highest in shopping streets and town centre roads. More recently, Cho et al. (2009) also showed that compact and mixed land-use areas were positively related with (actual) accident risks for cyclists and pedestrians, and that higher perceived risks were found in compact neighbourhoods only (reduced perceived risks were reported for highly mixed land uses). Last but not least, Thomas et al. (2011) found that the accident numbers and risks for cyclists were the highest in the central parts of the Antwerp's urban region, compared to the periphery. As shown in Chapter 2, the results tend however to be quite different when accounting for the severity of the accident. They may show reduced risks of being killed or seriously injured in urbanized areas, owing to a lower differential between slow and fast modes of transport.

In this chapter, the proximity of the accident/control to the town centre is measured by the network distance between the Brussels' town hall and each accident/control. It is here used as a proxy of the pedestrian activity, as well as for the accessibility to jobs/facilities (e.g. services, shops, restaurants, etc.) and population.

Major roads

The risk of having an accident is expected to be higher for cyclists riding on (or close to) major roads, i.e. on high-capacity and high-speed roads interconnecting large towns and attractive places or activities within/outside an urban area (see e.g. Klop and Khattak, 1999; Kim et al., 2007; Eluru et al., 2008). Such roads are indeed characterised by a design, traffic conditions and rules that are generally car-oriented but not in favour of a comfortable and safe cycling trip. For instance, the high speed limits (only achieved/exceeded by the motorists during off-peak hours) as well as the dense and conflicting traffic conditions (during peak hours) are factors that increase the probability of fatal and non-fatal accidents for cyclists, respectively (see Section 5.3.2.2 for further information on the safety effect of traffic conditions).

Data about major roads are provided by the Brussels UrbIS database (BRIC) and include several categories of roads (i.e. motorways, metropolitan roads, secondary/main roads and inter-district connecting roads). Two variables are here defined: one identifies the presence (or the absence) of a major road at the place of the accident/control, while the other computes the network distance to the closest intersection with a major road. In the latter case, it hence aims at measuring the influence of the proximity of major roads on the risk of bicycle accidents. It is assumed here that roads located in the neighbourhood of a major road have higher volumes of motorized traffic, and thus are more likely to exhibit a high risk of bicycle accident.

5.3.2.2 Traffic conditions

The number and risk of bicycle accidents are generally influenced by the traffic conditions (i.e. traffic composition, flows/volumes, etc.) observed at the time of the accident (see e.g. McClintock and Cleary, 1996; Klop and Khattak, 1999; Wang and Nihan, 2004; Hels and Orozova-Bekkevold, 2007; Kim et al., 2007; Eluru et al., 2008; Anderson, 2009). Such conditions spatially and temporally vary, depending on the type/design of road (major *vs* minor) and the time of the day (peak *vs* off-peak hours). In particular, during peak hours, a dense motorized traffic (or congestion) increases not only the number and the risk of non-fatal accidents for cyclists but also the perception of danger (Parkin et al., 2007; Hels and Orozova-Bekkevold, 2007; Møller and Hels, 2008), mainly because of the increased complexity of the traffic situation (e.g. high number of road users), the more aggressive driving behaviour (whatever the road user) and the restricted space left to the (passing) cyclists between the queuing vehicles (McClintock and Cleary, 1996; Li et al., 2007; Wang et al., 2009). It however decreases the risk of being seriously or fatally injured in a road accident, owing to a reduced differential between the speed of slow and fast transport modes (Klop and Khattak, 1999). More time is also given to drivers to react to conflicting situations and – as a corollary – to avoid accidents (Wang and Nihan, 2004; Wang et al., 2009). During off-peak hours, the opposite situation is observed. High vehicle speeds may be achieved (or exceeded) at these moments, which hence increases the risk of being seriously or fatally injured for cyclists (while reducing the probability of the other injury severities) (Klop and Khattak, 1999; Hels and Orozova-Bekkevold, 2007; Kim et al., 2007; Eluru et al., 2008). For instance, Kim et al. (2007) found a more than 11-fold increase in the probability of fatal injury as the estimated vehicle speeds pass 65 km/h.

The type of collision partner (e.g. pedestrian, car user, etc.) also plays a key role in the severity of the accident. Depending on their speed, dimension and weight, they may lead to different injury severities (the highest these three factors, the highest the injury severity). Disregarding the underreporting issue, motorised vehicles – and more particularly cars – generally account for the largest share of vehicles colliding with cyclists and often cause most of injuries for these latter (ERSO, 2006; Chong et al., 2010; Loo and Tsui, 2010). Lorries, buses, vans and sports utility vehicles are more frequently involved in serious and fatal cycling accidents, especially in urban areas where vulnerable road users and motorized vehicles interact (McCarthy and Gilbert, 1996; ERSO, 2006; Kim et al., 2007; Eluru et al., 2008; BRSI, 2009a; Pei et al., 2010; Yan et al., 2011). The wide vehicle dimensions (e.g. higher hood, reduced visibility for other road users being in the close vicinity, and larger blind spots than cars) combined with the relatively high speeds and heavier vehicle masses are some of the main factors explaining such severe injuries (ERSO, 2006; Kim et al., 2007; Eluru et al., 2008; Pei et al., 2010; Pai, 2011). In particular, these often occur when the cyclist is in the blind spot of a lorry turning right at a junction or when he/she is blown off his/her bicycle by a close lorry. Moreover, recent findings (Walker, 2007; Parkin and Meyers, 2010) show that drivers of buses and heavy good vehicles often leave narrow safety margins when overtaking cyclists due to their large dimensions (length) and poor acceleration (which may hence create close proximities and conflicts). They are also more prone to fatigue and stress since most of them ride within a commercial context, i.e. they have a planning and specific objectives to achieve (Boufous and Williamson, 2006; Brodie et al., 2009; Pei et al., 2010). Finally, the accidents of cyclists with other non-motorized road users – although less frequent – are not inconsequential in terms of injury, especially for pedestrians (McClintock and Cleary, 1996; Graw and König, 2002; Hels and Orozova-Bekkevold, 2007; Chong et al., 2010).

Data on motorized traffic volume were modelled by STRATEC for the year 2006 and are provided by the Brussels' Institute for Environmental Management (IBGE-BIM). They are expressed in terms of private car equivalent units and are measured for specific vehicle types, road links and time intervals. However, traffic modelling is computed for major roads only. In order to account for the traffic volume on minor roads as well as for the eventual bias generated by traffic modelling, a categorisation of the data into 5 classes is carried out based on the methodology of Natural Breaks (class 1: very low traffic level; …; class 5: very high traffic level). Within the framework of this chapter, only three vehicle types (i.e. car, van and lorry traffic) and three time intervals (i.e. 8:00 a.m.–8:59 a.m., 5:00 p.m.–5:59 p.m., and 6:00 a.m.–10:59 p.m.) are considered. Also, control is made of the street side of the accident / control when assigning traffic levels to these latter, except at intersections (where only the maximum traffic level is considered). If separated cycle facilities (i.e. uni- or bi-directional) are built along road links, cyclists are assumed to ride off-road and are hence supposed to be unaffected by the motorised traffic volume, except at intersections where the cycle facilities generally cross the road without any physical segregation. It hence leads us to assign a null traffic volume for accidents / controls occurring outside intersections and reported on the street side of such separated cycle facilities.

5.3.2.3 Environmental risk factors

Slopes / gradients

Road sections with steep slopes are expected to increase the risk of cycling accident. Straight gradients (downgrade) indeed reduce the control of the bicycle and lead to greater braking distances. They are also often associated with a greater number of curves on the road, which may limit the visibility of both motorists and cyclists (Klop and Khattak, 1999; Kim et al., 2007).

A Digital Elevation Model (DEM) − which is here represented as a raster of height values assigned to 90 × 90m pixels − is obtained from EROS Center (Earth Resources Observation and Science Center) and is used to compute the gradients at the place of the accident/control. These gradients correspond to the maximum slope (in degrees) computed between the pixel of the accident / control and the closest neighbouring pixels (Queen Contiguity, 1st order). Such a measure has however the disadvantage of not being computed along the road network and it hence does not account for infrastructures that are not constrained by the topography (e.g. bridges or tunnels).

Proximity to green areas

At our knowledge, there is no literature considering the effect of the proximity of green areas on the risk (and the number) of cycling accidents, and no trivial hypothesis could be made about such an effect. Although the close vicinity of green areas may increase the risk of accidents for cyclists in autumn (due to a skidding road surface caused by humid leave heaps), the effect of such a proximity is more difficult to assess as regards the intensity of recreational activities occurring in (or close to) green areas. On the one hand, these recreational activities can increase the risk of having a cycling accident (involving e.g. inattentive children that played in the neighbourhood of the green areas), whereas on the other hand this could make passing motorists more careful about potential conflicts with cyclists and pedestrians (that are in great numbers in the close vicinity of green areas, hence creating a ‘safety in numbers’ effect).

Data on green areas (i.e. parks, playgrounds, forests and woods) are provided by the Brussels UrbIS database (BRIC). Euclidean distances are first computed between each accident/control and the border of the closest green area. Second, a binary variable is created on the basis of the Euclidean distances in order to indicate the presence (or the absence) of a green area within some specified buffer distance from the place of the accident/control (10, 20, 30, 40, or 50m).

5.3.2.4 Interaction variables and intersect analyses

Most of the previous risk factors are combined/intersected and introduced in the models through trial and error processes. Crossing two risk factors may be advantageous since it may improve the inference (e.g. by obtaining a significant interaction variable, whereas the two risk factors taken separately may appear as being insignificant) and the validity of the models (owing to e.g. a reduced multicollinearity, heteroscedasticity, etc.). In the tables reporting the results of the final models, interaction variables are noted using the following notation: ‘[Risk factor 1] & [Risk factor 2]’. For instance, the interaction variable ‘Bridge & no cycle facility’ represents the bridges for which there is no cycle facility. Of note is also the fact that the influence of (nearby) traffic volume, parking areas, and tram tracks is disregarded when accidents/controls occur on separated cycle facilities and outside junctions, because the cyclists ride on cycle facilities that are physically separated from the road (and its attendant motorised traffic, etc.).

5.3.2.5 Ignored risk factors

Some infrastructure factors are deliberately ignored due to frequent modifications/treatments (for instance, a large number of advanced stop zones for cyclists were implemented at intersections during our period of study). Also, human and vehicle-related factors (as well as some of the environmental factors) are disregarded since: (1) they are not available for controls; (2) they are expected to be erroneously described for some accidents; (3) we deliberately focussed on the effect of modifiable risk factors only.

5.4 Results[4]

Descriptive statistics are reported in Section 5.4.1; they explore the relationships between the risk factors and the occurrence of bicycle accidents. Overall model evaluation and diagnostics are succinctly presented in Section 5.4.2, with the aim to motivate the selection of the final (autologistic) model. Results are discussed in Section 5.4.3, before illustrating the interest of such a modelling approach using predictions for a specific road trajectory in the Brussels' town centre (Section 5.4.4).

5.4.1 Bivariate associations

Chi-Square adjusted tests and Fisher's exact tests for independence confirm our expectations (Appendix D.2, discrete data) and show that there is a significant relationship between the presence of a number of risk factors and the occurrence of a bicycle accident. It also indicates that the probability of having a cycling accident is higher in intersections (i.e. right-of-way, yield/stop, traffic light and roundabout), on bridges, tram tracks (all types), cycle facilities (especially on unidirectional separated, marked, suggested, and bus-bicycle lanes), cycle facilities built next to parking areas (especially the unidirectional separated lanes), major roads or roads with low to very high volumes of motorized traffic (cars, vans and trucks), and when 1-10 garages are present within a 100m network distance. At the opposite, the probability of having an accident is reduced when cycling in contraflow streets or streets characterized by a very low motorized traffic volume (car, van or truck), by longitudinal or perpendicular-angled parking, by a number of garages ranging between 21 and 40 within a 100 network distance (or by at least 1 garage within 10 or 50m), and when cycling outside intersections, tram tracks or cycle facilities.

As regards the continuous data, Appendix D.3 shows that the probability of having an accident for the cyclist is lower for large garage lengths (within a 100m network distance), whereas it increases with the complexity of the location (whatever the bandwidth) and nearby the town centre, crossroads,

[4] All descriptive statistics and models performed within a frequentist framework were run in SAS Enterprise Guide 4.2 and R 2.12.1., while Bayesian statistics were computed in WinBUGS from R, by using the R2WinBUGS package (Sturtz et al., 2005). WinBUGS is the windows-based version of the BUGS software (BUGS: 'Bayesian inference Using Gibbs Sampling') and makes possible the use of several MCMC methods for analysis of hierarchical Bayesian models (Lunn et al., 2000; Spiegelhalter et al., 2003). Convergence diagnostics and output analyses were performed using the CODA package in R (Best et al., 1995; Plummer et al., 2006).

discontinuities in the bicycle network, major roads, parking areas, public transport stops, administrations, schools (all types, except primary and secondary schools), shopping centres, cultural buildings, religious buildings (almost all types), police buildings, hospitals and embassies. Non-parametric Wilcoxon Rank-Sum tests support such findings by suggesting that these risk factors significantly differ (in their median values) between the accidents and the controls.

5.4.2 Model diagnostics and selection

5.4.2.1 Accident risk modelling

Logistic modelling is first performed within a frequentist framework (see Appendix D.4) in order to identify which are the most significant risk factors and to get initial values for the parameters of the hierarchical Bayesian models. In this first step, special attention is paid to the level of multicollinearity (VIF_{max}=1.22) and to the presence of heteroskedasticity (Huber-White correction was here applied), as well as to the goodness-of-fit and effectiveness of the model (compared to an intercept-only model). Goodness-of-fit statistics and inferential statistical tests are reported in Appendix D.5 and show that the logistic model fits the data well (LL = −1063.1; HL test statistic = 14.1) and is more effective than the null model (LR test statistic = 883.4). The measures of association and misclassification also indicate that the model correctly predicts higher probabilities for accidents compared to controls (c = 0.83; D_{xy} = 0.66) and misclassifies only 14% of the observations (when setting the cut-off point of classification at 0.5). Wald Chi-Square statistics finally show that most of the parameters included in the logistic model are significant at the 95% level.

In a second step, the same logistic model is performed within a Bayesian framework. Table 5.1 (left columns) shows that the (posterior) values of the parameter estimates are very close to these computed within the frequentist framework (in Appendix D.4). Note that for all models, distance-based variables (i.e. distance to shopping centres or regional administrations) are exponentially transformed in order to improve the model fit[5].

[5] It hence suggests that the influence of the proximity of shopping centres and regional administrations on the occurrence of accidents adopts a negative exponential form ($e^{-0.001.x}$, where x is the distance – expressed in km – between the accident/control and the shopping centre/regional administration).

5.4.2.2 Autologistic and autoregressive accident risk modelling

Residual spatial autocorrelation is detected using Moran's I index ($I = 0.27$). This suggests that spatial dependence initially observed for the accidents (through join-count test statistics) is not fully taken into account by the selected risk factors. Autologistic and random effect specifications (i.e. with correlated and/or uncorrelated heterogeneity) are then implemented within a hierarchical Bayesian framework in order to deal with the presence of spatial autocorrelation in the model. Table 5.2 lists the best models (among approximately 100 models, with various specifications and complexities) and makes the comparison with the corresponding null model specifications. The autologistic formulation turned out to be the best (model IX). It not only provided evidence for robust convergence (through convergence diagnostics; see Appendix D.6), but also resulted in the smallest DIC value (2118.1) and in an improvement over the null models. On the contrary, the specifications with random effects did not succeed in converging and provided insignificant parameter estimates for both the uncorrelated and correlated random effects: the parameter estimates – or, more exactly, the Bayesian posterior mean estimates – were in the order of 1.10^{-5} and 1.10^{-20}, respectively. Several model specifications incorporating different sets of risk factors and random effects and using different burn-in periods, thinning intervals, or prior and hyper-prior distributions were tested here, but still without success. This hence suggests that both the uncorrelated and correlated random effects are weak (e.g. due to the inclusion of appropriate risk factors). However, the fact that an autologistic model previously converged may also suggest that the binary spatial weight matrix used for the ICAR model is a too simplistic form and is not convenient to capture the spatial – or rather the 'network' – autocorrelation. A spatial weight matrix such as this defined for the autologistic model (i.e. based on a decay function) seems to be a more appropriate form to account for the presence of spatial autocorrelation and probably explains why the autocovariate is significant in the autologistic specification (whereas random effects are insignificant in the ICAR model).

5.4.3 Discussion of the results of the autologistic model

Table 5.1 (right columns) presents the results of the autologistic model. It shows that almost all risk factors are significant at 95% and that the MAPE is quite small, indicating a low misclassification under the fitted model. Only infrastructure- and traffic-related risk factors are retained in the model.

Table 5.1: Logistic (non-spatial) and auto-logistic models (spatial) – Results from the Bayesian framework

Variables	Logistic model						Autologistic model					
	Mean	SD	MC error	CI 2.50%	CI 97.50%	OR	Mean	SD	MC error	CI 2.50%	CI 97.50%	OR (OR_{100m})
Intercept[a]	-2.29***	0.09	0.001	-2.47	-2.12	0.10	-2.29***	0.09	0.001	-2.46	-2.12	0.10
Autocovariate variable	-	-	-	-	-		2.15***	0.14	0.001	1.89	2.42	8.61
Infrastructure												
Complexity index												
Bandwidth = 10m	0.15***	0.01	0.000	0.13	0.17	1.16	-	-	-	-	-	-
Bandwidth = 40m	-	-	-	-	-	-	0.02***	0.00	0.000	0.01	0.02	1.02 (*4.79*)
Bridge & no cycle facility	0.86	0.58	0.006	-0.29	2.00	2.37	0.88	0.59	0.005	-0.26	2.03	2.42
Contraflow cycling & no crossroad	-0.69*	0.35	0.003	-1.42	-0.05	0.50	-0.89**	0.36	0.003	-1.64	-0.23	0.41
Cycle facility & crossroad												
Fac.1 (unidir.) & Crossr.1 (yield/stop)	2.25**	0.92	0.009	0.63	4.27	9.53	2.02**	0.90	0.008	0.44	3.99	7.56
Fac.2 (bidir.) & Crossr.1 (yield/stop)	2.88**	1.38	0.013	0.66	6.02	17.78	3.36***	1.38	0.012	1.15	6.56	28.85
Fac.3 (mark.) & Crossr.3 (traff. light)	1.96**	0.94	0.009	0.32	4.01	7.10	1.85*	0.91	0.007	0.25	3.79	6.35
Fac.3 (mark.) & Crossr.4 (round.)	2.76*	1.52	0.013	0.18	6.13	15.83	2.83*	1.56	0.013	0.13	6.22	16.91
Fac.4 (sugg.) & Crossr.2 (right-of-w.)	3.13**	1.42	0.012	0.87	6.46	22.90	3.74***	1.37	0.011	1.60	7.05	42.22
Fac.0 (no facility) & Crossr.4 (round.)	1.02***	0.30	0.003	0.43	1.61	2.78	0.67*	0.32	0.002	0.03	1.30	1.96
Fac.3 (mark.) & Crossr.0 (no crossr.)	0.73*	0.33	0.003	0.06	1.35	2.07	-	-	-	-	-	-
Tram tracks												
Class 1 (crossing tram tracks)	0.86*	0.44	0.004	0.01	1.75	2.37	1.16**	0.46	0.004	0.29	2.09	3.20
Class 2 (crossable reserved lanes)	0.83**	0.33	0.003	0.17	1.47	2.30	-	-	-	-	-	-
Class 3 (on-road tracks)	1.06***	0.23	0.002	0.60	1.51	2.87	0.82***	0.23	0.002	0.36	1.28	2.27

continued on next page

continued

Variables	Logistic model						Autologistic model					
	Mean	SD	MC error	CI 2.50%	CI 97.50%	OR	Mean	SD	MC error	CI 2.50%	CI 97.50%	OR (OR_{100m})
Number of garages (d ≤100m)												
Range 0 (no garage)	-0.61*	0.28	0.003	-1.18	-0.07	0.54	-0.60*	0.28	0.002	-1.17	-0.07	0.55
Distance public administration[b]												
Public administration 2 (regional)	1.08***	0.22	0.002	0.65	1.52	2.95	-	-	-	-	-	-
Distance shopping center[b]	-	-	-	-	-	-	0.86***	0.24	0.002	0.38	1.33	2.36 (*2.17*)
Proximity parking-cycle facility												
Parking & Facility 1 (unidir.)	1.28**	0.45	0.004	0.37	2.14	3.59	1.15*	0.48	0.004	0.18	2.08	3.16
Parking & Facility 2 (bidir.)	2.07*	1.16	0.011	-0.22	4.40	7.95	1.76	1.30	0.011	-0.88	4.27	5.78
Traffic												
Van & truck traffic (6am-10:59pm)												
Class 2 (low)	1.01***	0.15	0.001	0.71	1.30	2.73	0.92***	0.15	0.001	0.64	1.21	2.52
Class 3 (moderate)	1.32***	0.16	0.001	1.01	1.63	3.75	1.20***	0.16	0.001	0.89	1.51	3.32
Class 4 (high)	1.24***	0.22	0.002	0.80	1.68	3.46	1.26***	0.22	0.002	0.82	1.70	3.53
Class 5 (very high)	2.60***	0.35	0.003	1.93	3.29	13.46	2.13***	0.36	0.003	1.43	2.84	8.38
Deviance	2149***	6.92	0.060	2137	2164	-	2097***	6.70	0.052	2086	2112	-
MAPE	0.21***	0.00	0.000	0.20	0.22	-	0.21***	0.00	0.000	0.20	0.21	-
MSPE	0.11***	0.00	0.000	0.11	0.11	-	0.10***	0.00	0.000	0.10	0.11	-

*** Significant at 99.9%; ** Significant at 99%; * Significant at 95%

[a] Intercept value resulting from centering

[b] Exponentially transformed variables ($e^{-0.001.x}$)

OR: Odds Ratio; OR_{100m}: Odds Ratio for a 100m increase (rather than 1m)

CI: credible interval

Interaction variables:

Bridge & no cycle facility: Bridge = 1 and Cycle facility = 0

Contraflow cycling & no crossroad: Contraflow cycling = 1 and Crossroad = 0

Van & truck traffic (6am-10:59pm): maximum class value of van & truck traffic

Cycle facility & crossroad: cycle facility = X, crossroad = Y (X=1, ...5; Y=1, ...6)

Table 5.2: Model-building specifications and model comparison

Model ID	Model	Iterations	Burn-in	Thin	p_D	DIC	Converged?
I	Null model	15000	10000	1	1.00	3011.53	Yes
II	Null model + UH	250000	50000	100	25.10	3016.88	No
III	Null model + Autocov	50000	10000	10	2.00	2572.26	Yes
IV	Null model + UH + Autocov	36000	9000	60	3.27	2572.3	No
V	Null model + CH	-	-	-	-	-	No
VI	Null model + UH + CH	-	-	-	-	-	No
VII	**Fixed effects only**	**50000**	**10000**	**10**	**22.21**	**2171.15**	**Yes**
VIII	Fixed effects + UH	35000	10000	5	27.02	2171.23	No
IX	**Fixed effects + Autocov**	**20000**	**10000**	**2**	**21.08**	**2118.08**	**Yes**
X	Fixed effects + UH + Autocov	20000	12500	10	22.60	2118.64	No
XI	Fixed effects + CH	-	-	-	-	-	No
XII	Fixed effects + UH + CH	-	-	-	-	-	No

Number of Markov Chains = 3

Null model: Model with intercept only (no risk factor included)

UH: Uncorrelated Heterogeneity (unstructured errors); CH: Correlated Heterogeneity (spatially structured errors, ICAR)

Autocov: Autocovariate; p_D: effective number of parameters; DIC: Deviance Information Criterion

Converged?: 'Yes': the model converged to a perfect equilibrium; 'No': the model did not converge to a perfect equilibrium, but the trace plots suggest it almost reached it.

In **bold**: selected models (VII = Logistic model; IX = Auto-logistic model)

Interestingly, variables referring to the gradients, the discontinuities in the bicycle network and the traffic-calming measures (i.e. pedestrian, residential and 30 km/h areas) are not included here. Although the topography is far from being flat in Brussels (especially in the southern part of the region), the gradients are not as steep as they are in some Walloon municipalities and do not seem to significantly increase the risk of having a cycling accident. The fact that few severe injuries are reported in Brussels probably confirms such a statement (given that steep slopes are generally more likely to increase the risk of injury severity, rather than the global risk). Concerning the discontinuities, these were removed from the model due to collinearity issues with some other risk factors (e.g. with the intersections, since most of the discontinuities are here located). Finally, traffic-calming areas led to insignificant parameters for most of the model specifications (except for pedestrian areas, in some cases). This reflects the lack of efficiency of such areas in reducing the accident risk, which is probably due to the current generalization of 30 km/h areas in Brussels and the little respect motorists have about speed limits. In 2007, about 77% and 45% of motorists driving in the BCR indeed committed an offence of more than 1 km/h and more than 10 km/h (respectively) in 30 km/h areas (BRSI, 2009b). Note that, in the next subsections, hypothetical explanations are provided for each risk factor on the basis of results and observations derived from the grey literature (especially BRSI, 2006, 2009a, 2009b) and from the review of the literature conducted in Section 5.3.2.

5.4.3.1 Infrastructure-related risk factors

Among the infrastructure-related variables, the complexity index has the largest effect on the risk of having a cycling accident. It accounts for about 30% of the explanation of the accident risk, whatever the location on the network (with a maximum of 92% for streets with at least one garage and where contraflow cycling is not permitted). As mentioned before, cyclists as well as other road users are faced with a large number of information at the same time at locations with an increased complexity (e.g. due to a high number of road legs, signs, road users, etc.). Driver errors are hence more likely to occur at such locations. This suggests that the complexity index computed here somewhat accounts for the driver behaviour (and hence not only the infrastructure-related aspect) in capturing the driver errors that could result from the reduced legibility of the urban streetscape.

Although significant at 93%, the parameter estimate corresponding to the bridges without cycle facilities suggests that the risk of bicycle accident increases at such locations. The sudden change in terms of road width (i.e. narrow space) and visibility (which is low due to the curving of the bridge) is expected to be at the root of such an increased risk, especially if no dedicated facility is built for cyclists on the bridge. If well-kept and designed, such a cycle facility could probably outweigh (or at least reduce) the risks generated by the low long-distance visibility and the narrowing of the road space.

Contrary to popular belief, we here show that contraflow cycling reduces the risk of having an accident for cyclists. Such a lower risk might result from a 'risk compensation effect', i.e. from the fact that drivers may tend to behave in a more cautious way due to an increased perceived risk in streets where contraflow cycling is permitted. At the opposite, drivers may tend to behave less carefully in places where they feel safer. Interestingly, the fact that intersections are here excluded from the definition of streets with contraflow cycling indicates that motorists entering into such streets may be surprised to be in front of (exiting) cyclists (probably because they do not already behave in a cautious way since they just enter into a contraflow street).

Regarding cycle facilities, the results are in line with the literature (see Chapter 4 for a literature review) and indicate that some of these facilities lead to an increased risk of having a bicycle accident when associated with a specific type of intersection. In particular, right-of-way intersections equipped with suggested cycle lanes lead to the highest accident risk for cyclists, probably because of the non-respect of the right-of-way by motorists (BRSI, 2009a) and the very discontinuous character of the facility (i.e. chevrons and bicycle logos only, instead of a 'continuous' lane or path). According to accident data registered in Brussels for the period 2006-2008 (DGSEI), collision partners indeed did not give way to the cyclist in about 59% of the accidents (whereas cyclists were held responsible for not giving way in about 10% of the accidents). Moreover, the discontinuous character of suggested cycle lanes possibly makes the cyclists less 'visible/expectable' for motorists approaching a right-of-way intersection, especially when compared to segregated or marked cycle facilities (that have a more continuous character). Yield/stop intersections with separated cycle lanes also seem to carry a danger, especially when the cyclist rides on a bidirectional facility in the opposite direction of the (parallel) traffic. The reasons are probably twofold: on the one hand, cyclists often have an ill-founded feeling of safety caused by the physical segregation of the facility, while on the other hand motorists often have an inappropriate visual search pattern (i.e. they often look at one direction only) and do not expect to cross a cyclist coming from an opposite direction (BRSI, 2006). It seems that the same accident mechanisms also apply to the cycling accidents at yield/stop intersections equipped with unidirectional separated lanes, where the cyclists sometimes ride in the wrong way (i.e. not permitted by law) (*ibid.*). Given that such facilities are frequently built on either side of multi-lane and divided roads, we assumed that – in this case – the cyclist was often deterred to cross the (wide and busy) road in order to be in the right way. As expected, high accident risks were also observed for cyclists riding on marked cycle lanes built in roundabouts (outer lane). In such a context, collisions often occur when the motorist leaves or enters into a roundabout and cuts in on the cyclist riding on the marked facility. Such a design even leads to a higher accident risk for cyclists compared to roundabouts without any cycle facility (where the cyclist is merged into the stream of motorized traffic). Intersections equipped with traffic lights and marked cycle lanes are also found to increase the risk of accident for cyclists. Such an increased risk is probably due to motorists turning to an adjacent road and cutting in on the (straight) cyclist's trajectory on the marked facility. This may also be explained by the fact that cycle lanes are generally designed in such a way that they position cyclists in the blind spots of the (large) motorised vehicles at signalised intersections. However, it is worth noting that accident risk is here lower compared with the above mentioned designs (it is about 7 times less risky than right-of-way intersections equipped with suggested cycle lanes). This is probably the result of a reduced number of conflicting movements and lower vehicle speeds at signalized intersections. Also, the presence of advanced stop zones for cyclists is expected to mitigate the accident risk at signalized intersections. Such zones are quite frequent here and are often used in conjunction with marked lanes. They not only put the cyclists into the view of motorists (and outside blind spots of cars and large vehicles), but also allow cyclists preparing to turn to take up a proper position on the road. More generally and disregarding the type of facility or intersection, it is not uncommon in Brussels that cycle facilities abruptly stop at intersections, providing no dedicated/safe room for the cyclist within the motorized traffic and hence increasing the probability of having an accident here. At some intersections, inappropriately designed and/or poorly maintained cycle facilities may also lead to confusing situations where it is not easy to determine which road user (cyclist or motorist) has to give way.

The close vicinity ($\leq$ 0.8m) between separated cycle lanes (both types) and parking facilities is also identified here as being a significant risk factor. Cyclists riding on such separated lanes and alongside close parked vehicles may indeed run into (suddenly) opened car doors. Also, the presence of parked vehicles generates a (close) pedestrian activity that may sometimes occur on the adjacent cycle lane (due e.g. to the absence of sidewalk, non-respect of the cycle facility, etc.) and may potentially lead to an accident. This is all the more true as, in Brussels, the joint presence of parked vehicles and separated cycle lanes is frequently observed alongside major roads, characterised by close attractive activities (e.g. business and industrial zones, residences, parks, etc.).

Similarly, Table 5.1 suggests that the presence of garage/parking driveway (within a 100m network distance) increases the risk of having an accident while cycling. This result may be explained by the fact that motorists leaving or entering into a garage/parking driveway may collide with cyclists riding straight ahead on the road (*ibid.*). Interestingly, Appendix D.2 also suggests that the risk is strongly increased in locations where few (1-10) and many garages (> 50) are observed within a 100m network distance. It is here assumed that a risk compensation effect applies, in the sense that cyclists are probably more cautious when riding in streets characterized by many garages/driveways than in streets where garages/driveways are less (or very) frequent. In the case where few garages are present (1-10), it is assumed that cycling accidents are caused by a 'surprise effect'. At the opposite, in a street where garages are quite frequent (> 50), accidents might be caused by the fact that cyclists are accustomed (after some period of time) to riding along garages and may take less care in spotting motorists leaving garages. Another assumption could be that cyclists are faced with a large number of information (i.e. the numerous garages) and, then, may have a reduced ability to make decisions in a very short time while cycling.

Concerning tram tracks, our findings indicate that the presence of on-road tracks and tram (tracks) crossings significantly increase the risk of having a bicycle accident. As suggested in Chapter 4, cyclists may get stuck in tram tracks, resulting in a loss of control of the bicycle (conducting to a fall, in some cases). It is also assumed here that the presence of on-road tracks forces the cyclist to ride on places that are not especially optimal for his/her own safety. For instance, he/she has to make the difficult choice between riding next to parked vehicles (exposing him/her to the opening of door cars) and riding between the tracks, i.e. in the middle of the road lane (exposing him/her to eventual aggressive drivers that are blocked behind and constrained to lower their driving speed).

Last but not least, the presence of a shopping centre or arcade in the close vicinity of the cyclist's trajectory is also associated with an increased risk of accident. An intense pedestrian and/or motorized activity is indeed commonly observed in the neighbourhood of shopping centres. This hence increases the number of potential conflicting situations, and then leads to a higher risk of accident for cyclists.

5.4.3.2 Traffic conditions

Among all traffic-related risk factors, those referring to the different levels of van and truck traffic (classes 2–5) provided the best improvement of the model fit. They are all highly significant and indicate that increasing levels of van and truck traffic are associated with higher accident risks. Whatever the type of road user, the complexity of the traffic context is indeed as much increased as the traffic is denser. The road legibility as well as the cognitive capacity of the road user are indeed reduced due to the presence of a large number of information to process in the streetscape (which reduces the ability to detect and carry out appropriate actions to control traffic hazards) (Elvik, 2006). Vans and trucks are also more prone to blind spot problems when turning and leave narrow safety margins to cyclists when overtaking (e.g. due to a wrong estimation of the overtaking time), which clearly increases the risk of accident (and injury severity) for cyclists. Furthermore, the large vehicle dimensions of vans and trucks may obstruct the field of vision of all neighbouring road users (i.e. cyclists, motorists, etc.) and – as a result – may lead to conflicting situations between these latter.

5.4.4 Predictions of the risk for a specific road trajectory: a tool for planners?

Given that the accident risk spatially varies along the network (as a function of the reported features/factors), it is rather tricky to assign here an order of importance to each risk factor. Rather, mapping the above predicted risk of having a bicycle accident may be quite interesting since it not only validates the results of the autologistic model, but it also provides a useful tool for planners, decision makers and cyclists' advocacy groups. As an illustration, predictions are here computed for sampled points located every 10m along a specific road trajectory and are afterwards interpolated along this trajectory using the approximate spline curve method from SANET v.4 (bandwidth = 100m; cell width = 2m) (Figure 5.5). The road trajectory passes through the Brussels' European district (Schuman roundabout (numbered 1* in Figure 5.5), Rue de la Loi / Wetstraat (2*)) and nearby the Pentagon (CBD, Royal Palace and Park (3–4)) and Brussels' University (ULB–VUB (12–13)). This road trajectory is selected mainly because of its high variability in terms of the risk factors identified along the cyclist's route, but also because of the high bicyclist volumes, the international reputation (for foreign scientists) and the in-depth authors' knowledge of the route (as cyclist, pedestrian and motorist). Note finally that the bicyclist's direction (i.e. street side) and the building year of the infrastructures (i.e. tram tracks, cycling and parking facilities, etc.) are taken into account as much as possible when assigning the risk factors to the sampled points. In the present case, the most recent year – i.e. 2008 – is selected for computing the predictions. Moreover, it is assumed that the cyclists travelling on the selected trajectory respect the law by riding on the cycle facilities (when present).

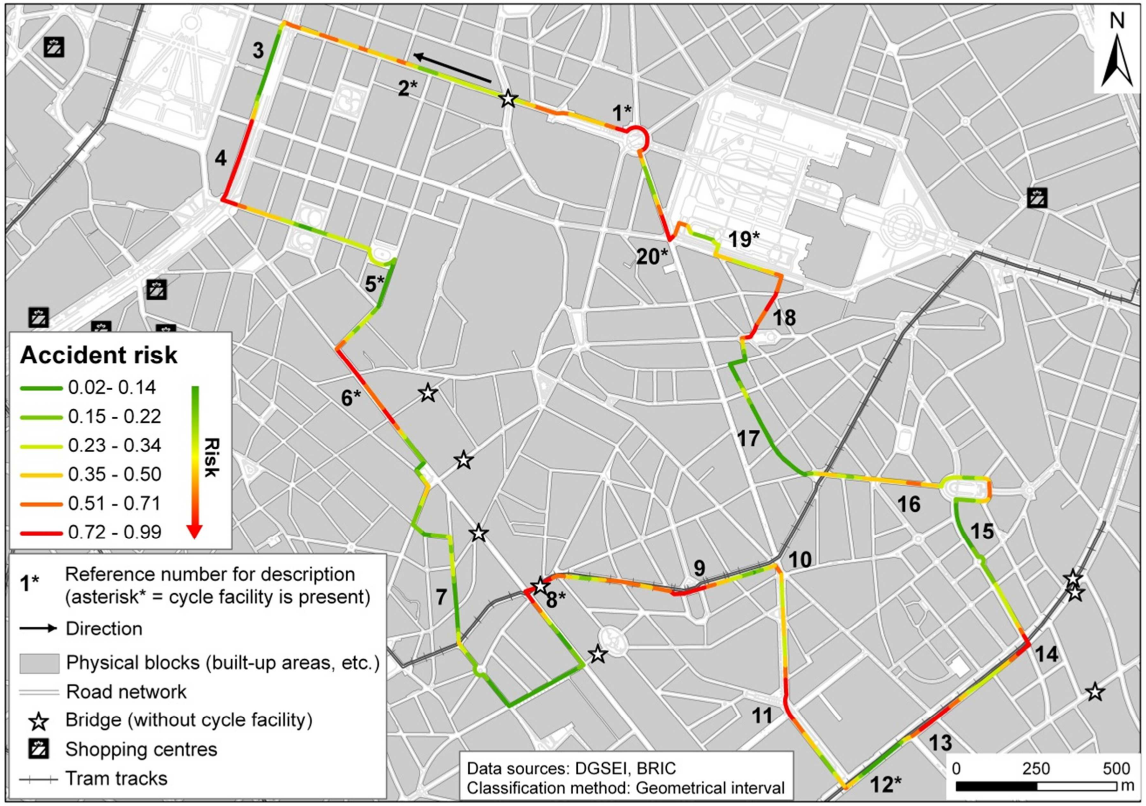


Figure 5.5: Map of the predicted risk of having a cycling accident, computed from parameter estimates of the autologistic model (road trajectory in Brussels)

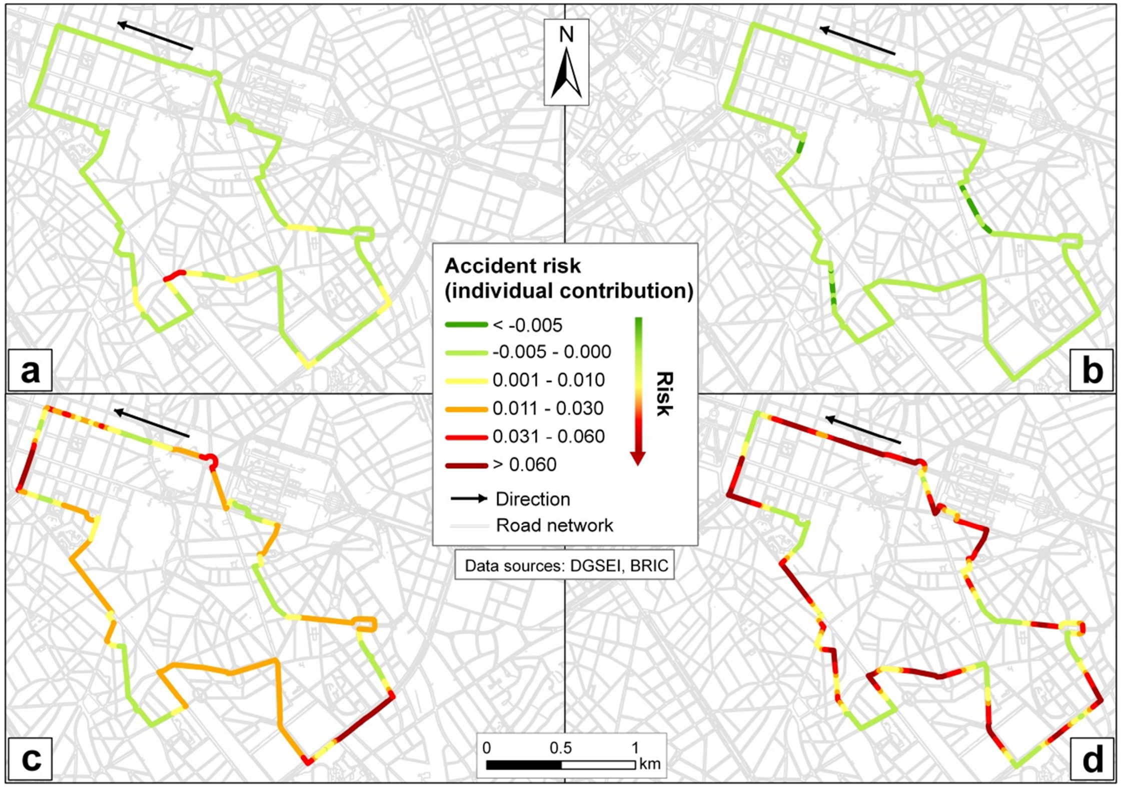


Figure 5.6: Predicted risk of having a cycling accident, separately computed for 4 risk factors: (a) tram tracks (on-road and crossings); (b) contraflow cycling (intersections are excluded); (c) van and truck traffic from 6 a.m. to 10:59 p.m. (all classes from 2 to 5); (d) autocovariate component.

Figure 5.5 identifies the most ‘risky’ parts of the trajectory, and hence the places where cyclists should be more careful when riding and/or where changes in the infrastructures might be performed in order to improve bicyclist’s safety. In particular, red-coloured links correspond to locations where the accident risk for cyclists is the highest, whereas green ones represent locations where risk is the lowest. Figure 5.6 also exhibits the individual contribution of 4 of the risk factors to the total risk of accident for cyclists. These referring to the streets with contraflow cycling and tram tracks contribute (positively or negatively) to the total risk of accident at a very local scale (since they are infrastructure-dependent), whereas van and truck traffic volumes and the autocovariate component have a more spatially loose effect along the trajectory. It is here assumed that such an autocovariate component captures the unobserved/unidentified risk factors (random effects) that are entirely specific to each accident location. Interestingly, both maps suggest that the risk of bicycle accident is higher for ‘complex’ intersections (i.e. those numbered 1^*, 8^*, 10–11, 15–16, 18, 20^*), roundabouts with marked cycle lanes (1^*), roads with on-road tram tracks and tram crossings (8^*– 9, 12^*, 14), as well as roads with dense van and truck traffic volumes (1^*, 2^*, 4, 6^*, 8^*, 11, 13–14, 18, 20^*). At the opposite, the lowest accident risks are mainly observed for streets located in residential wards (characterized by low van and truck traffic volumes (5^*, 9, 16)), where contraflow cycling is allowed (5^*, 7, 17), or where no garage is observed in the close vicinity (1^*, 5^*, 12^*, 19^*).

Predicting the risk of cycling accident on the entire network may also provide several important advantages over black spot methods (see Figure 5.4). Given that cycling accidents are strongly under-reported in Brussels, it is expected that modelling methods based on reported accident data (and using all the related ‘information’ on risk factors) would be here helpful in identifying the locations where unreported cycling accidents might have occurred. Such methods indeed exploit **all the available information** from the accident data set (and from all accidents) in order to compute a predicted risk of accident **for every point of the network**. On the contrary, black spot methods do not take advantage of using such information to infer locations where accidents might have been unreported (since they only describe/identify the spatial concentrations of registered accidents). As an illustration, comparing Figure 5.4 with Figure 5.5 for the same road trajectory shows that the risk of cycling accident is far from being negligible in points where there is actually no reported cycling accident but where they are yet expected to occur (due to e.g. a dense traffic, or the presence of tram tracks). For instance, locations 3 and 10 correspond to major intersections (i.e. with a dense motorised and pedestrian traffic) where there is no reported cycling accident but where it is quite doubtful that it is actually the case in view of the local traffic conditions. In line with our feelings, Figure 5.5

shows that the probability of having an accident is predicted to range between 51 and 71%, which suggests that cycling accidents might have been unreported here (which is to be expected).

Finally, black spot methods do not take into account the building year of the infrastructure, which could be an issue when working on several years (e.g. on a three-year period) since a location may be informed as being ‘dangerous’ whereas it could not be anymore the case after having carried out some important infrastructure changes during the period of study. More importantly, most of the black spot methods do not consider the traffic direction and may indicate both sides of a street as being dangerous for cyclists whereas most cycling accidents cluster on one side only. This is for example the case for the road link running from the bridge in Rue de la Loi / Wetstraat (between 1^* and 2^*) to location 2^*: Figure 5.4 designates it as a black spot for cyclists whereas the probability of having a cycling accident is quite low (< 34%) in the direction indicated by the arrow in Figure 5.5. Black spot methods hence fail to accurately identify the ‘dangerous’ street side and may lead to erroneous recommendations and decisions about infrastructures. At the opposite, the measurements over the predicted risk of having an accident (Figure 5.5) here take into account the building year of the infrastructure as well as the bicyclist’s direction, and – as a corollary – seem to be closer to the reality than the results obtained from black spot methods.

5.5 Conclusion

The main objective of this chapter was to identify which spatial variables (or risk factors) are significantly associated with the occurrence of cycling accidents in the Brussels-Capital Region. This chapter is original in many ways. Taking advantage of the recent research implemented in epidemiology, ecology and transport geography, this chapter opens up a new direction of research in traffic accident analysis by suggesting to use a case-control strategy to estimate a so-called ‘accident risk model’ at a micro-scale. In order to make possible the use of (auto-)logistic and conditional autoregressive modelling, a binary variable was constructed by adding controls (i.e. locations where there is no reported cycling accident) to the geocoded accident data set. Controls were then sampled along the Brussels’ bikeable road network as a function of the bicycle traffic (estimated using a gravity-based index) and excluding the locations where cycling accidents were reported by the police.

Although time-consuming, a rigorous digitization process was carried out at a micro-scale in order to collect GIS data on potential risk factors (i.e. expected to be associated with the occurrence of cycling accidents) for the whole Brussels’ network. A modelling process was finally performed within a Bayesian framework to highlight the most significant factors influencing the risk of cycling accident, as well as to identify the 'dangerous' locations for cyclists by mapping the predicted risk of accident along a specific road trajectory of the Brussels' network (Figure 5.5). Such predicted values of the risk then offer for planners and decision makers a new tool that accurately locates the places/streets at high risk of accident for cyclists (especially by accounting for the bicyclist's direction and the building year of infrastructures). It hence yields useful information to help cyclists choosing the safest route for their journeys (see Chapter 6 for further details). Interestingly, our modelling approach also has the advantage to exploit all the available information of the accident data set (and about risk factors) to pinpoint the locations where cycling accidents might have been underreported.

Methodologically, our results showed that the autologistic model turned out to be the best specification and conducted to the best results, whereas specifications incorporating random effects (e.g. ICAR model) did not succeed in converging and provided insignificant parameter estimates for both random effects (which either indicates that appropriate risk factors are included in the model, or suggests that the binary spatial weight matrix used for the ICAR model is a too simplistic form and is not convenient to capture the spatial autocorrelation).

From a planner's point of view, a plethora of results is obtained throughout this chapter. Contrary to motorists' beliefs, our results first show that streets where contraflow cycling is permitted reduce the accident risk, which hence supports a wider implementation of such streets in Brussels (although great care should be taken when designing these at intersections). At the opposite, our findings also indicate that most of the other risk factors increase the accident risks for cyclists. In line with the literature in traffic accident research (see e.g. Räsänen and Summala, 1998; Autlman-Hall and Hall, 1998; Aultman-Hall and Kaltenecker, 1999), results first suggest that cycle facilities significantly increase the risk of accident when they are combined with a specific type of intersection. Suggested cycle lanes crossing right-of-way intersections exhibit the highest accident risk for cyclists, probably because of the non-respect of the right-of-way and/or because of the discontinuous character of the facility (which makes it less visible). Uni- and bi-directional separated cycle lanes built at yield/stop intersections also carry a danger for cyclists, since motorists may adopt here an inappropriate visual search pattern (i.e. they look at one direction only) while cyclists may have an ill-founded feeling of safety caused by the physical separation from the road. Roundabouts and signalized intersections equipped with marked cycle lanes increase the risk of accident of cyclists as well. Previous research (focussed on accident mechanisms) suggests that accidents are frequently caused by motorists leaving/entering into the intersection and cutting in on the trajectory of the cyclist riding on the cycle lane. As regards the signalised intersections, such an increased risk may be explained by the fact that marked cycle lanes are designed in such a way that cyclists are positioned in the blind spot of trucks and vans. Regardless of the effect of the cycle facilities, it is also well-known that intersections are 'hot spots' of accidents for cyclists (as well as for all road users) given that the number of potential conflict points is far higher compared to the rest of the network (see e.g. Wang and Nihan, 2004; Reynolds et al., 2009; Haque et al., 2010). Moreover, as suggested throughout this chapter, intersections may be considered as 'complex' locations since road users must handle here a large number of information at the same time. Driver errors are hence more likely to occur at these places than anywhere else (Elvik, 2006).

Second, our results provide robust evidence for an increased risk of accident for cyclists who ride on bridges or in the close proximity of garages, parked vehicles (combined with separated cycle facilities) and shopping centres. As regards bridges, sudden changes in infrastructures (e.g. narrower space) and road conditions (e.g. bridges are more prone to ice development) may explain such a higher risk. As expected, the presence of garage/parking driveways and parked vehicles close to separated cycle facilities also significantly increase the risk of running an accident when cycling. Vehicles leaving/entering into garages and cutting in on the cyclists' trajectory (in the former case) as well as opened car doors and/or pedestrian activity occurring on the cycle facilities (latter case) may explain to some extent such an increased risk for cyclists.

Last but not least, this chapter reveals that an increased risk of cycling accident is significantly associated with the presence of on-road tram tracks in a street and with high levels of van and truck traffic. Cyclists indeed carry the danger of getting one of their wheels stuck in tracks, resulting in a loss of control of the bicycle and then possibly in a fall. Besides, our results show that streets with high levels of van/truck traffic are significantly associated with higher accident risks, which is probably explained by the fact that such streets generally correspond to major roads (i.e. interconnecting important places and designed to allow high traffic volumes as well as vehicles with large dimensions). Previous studies also frequently suggest that cyclists overtaking/riding along vans and trucks are more prone to be undetected by other road users, due to e.g. the higher likelihood to ride in blind spots of van/truck drivers. Similarly, the large dimensions of vans and trucks may hide cyclists and put these out of sight of other road users (e.g. car drivers). To our knowledge, this last hypothesis has however not been confirmed yet in the literature and would be worth testing in further research.

As mentioned before, these results serve as a basis for some of the safety-oriented recommendations approached in Chapter 6. Such recommendations are intended for policy makers and planners, with the aim to provide a sound scientific support for making bicycle use safer and, then, more common in Brussels.

Part IV: General conclusions and policy implications

Chapter 6

Conclusion

This thesis aimed at identifying the spatial factors that influence the use of the bicycle for commuting to work, as well as those that are associated with a reduced/increased risk of cycling accident. Complementarily to this general objective, it also had the intent to come up to policy makers and planners' expectations by providing further science-based knowledge on cycling. The use of the bicycle as a mode of transport indeed arouses the interest of policies oriented towards a sustainable development of the society, as it holds the potential to tackle a plethora of concerns related to the mobility, environment and public health. To achieve these goals, this thesis adopted a multidisciplinary approach and drew its inspiration from several scientific fields sharing more or less interest for the analysis of spatial data (e.g. quantitative geography, spatial econometrics, epidemiology, etc.).

From an *empirical point of view*, the objective of this thesis was two-fold. On the one hand, it focused on Belgium and aimed at investigating the relationship between cycle commuting and accident risks for cyclists, after which it identified the potential impact of a wide range of spatial factors on cycle commuting. In this latter case, special attention was paid to bicycle-specific factors (e.g. cycle facilities, hilliness), which turn out to be used in only a few works in mode choice research. As they are directly related to the use of the bicycle, they are expected to play a prominent role in explaining the spatial variation of cycle commuting in Belgium (even when controlling other confounding factors). On the other hand, the aim was to examine the spatial factors that are associated with the risk of being involved in a road accident when cycling along the Brussels' network (capital of Belgium). An initial point pattern analysis was also conducted beforehand in order to examine whether or not official accident databases neglect important information relative to unreported cycling accidents (e.g. as regards some specific risk factors). High-resolution factors related specifically to cycling accidents are here manually digitized into a GIS and then compiled in an exhaustive database. Several of these factors are – to our knowledge – considered for the first time in the literature on traffic accidents.

From a *methodological point of view*, this dissertation mostly aimed at accounting for a number of spatial effects associated with the data sets. Markedly, empirical studies in mode choice research rarely if ever attempted to correct biases resulting from the presence of spatial autocorrelation and heterogeneity in the models. Such 'aspatial' approaches then carry the danger to provide wrong policy recommendations if spatial data are included in the model. Within the framework of this thesis, and contrary to the vast body of literature, it was hence aimed to consider such spatial effects by performing appropriate statistical models. Of concern is also the fact that many studies in traffic accident research still assume a planar space as real world when attempting to pinpoint 'hot spots' of (cycle) accidents. Several studies indeed emphasized that it might lead to biased estimates. Taking advantage of recent advances in GIS, this thesis then devoted particular attention to the methods extended for network spaces and applied these to explore and compare the spatial patterns of cycling accidents officially registered by the police with those that are unregistered. As mentioned further, this latter approach provided further insights on the 'locational tendencies' of underreporting (i.e. *where* underreporting occurs) and on the bias it could bring throughout a modelling approach. Lastly, issues frequently stressed in the literature are also the lack of reliable data on the factors that influence the occurrence of cycling accidents, as well as on the trip characteristics of cyclists (exposure data) and accidents themselves (underreporting). Such issues often hamper to get in-depth insight on the actual risk of being involved in a road accident when cycling, except when surveys are conducted among the entire population of cyclists. Although time-consuming, these surveys indeed open the possibility of collecting both exposure and accident data and in turn allow estimating the accident risks. Such surveys however raise several questions about the way controls are selected, and then about their overall relevance in providing reliable parameter estimates. An innovative methodological framework, based on a rigorous sampling design of controls, was then proposed in this thesis to model the risk of cycling accident. Interestingly, it provided new directions of research for pinpointing 'risky' locations where (cycle) accidents occur along the road network.

This conclusive chapter is structured as follows. Section 6.1 summarizes the main findings of this thesis, after which Section 6.2 highlights the main implications this thesis has for planners and policy makers and Section 6.3 describes the limitations encountered throughout this thesis. It finally ends by providing perspectives for future research (Section 6.4) and some concluding words (Section 6.5).

6.1 Main findings

Throughout this thesis, the intent was to obtain sound results in order to enable planners and policy makers to have strong science-based support to encourage cycling and make it safer. To achieve this, special attention was paid to the methodological and data limitations reported in Chapter 1. These limitations referred to: (*i*) the lack of data; (*ii*) underreporting of cycling accidents; (*iii*) spatial data and attendant issues; (*iv*) network phenomena and planar assumption; (*v*) estimation of accident risks. They were all consistently taken into account – or at least monitored as regards their impacts – through the use of appropriate methods and exhaustive data collection. This section then first provides major conclusions as regards the way methodological and data limitations were tackled in this thesis (Section 6.1.1), after which it presents some of the main empirical results obtained by this way (Section 6.1.2).

6.1.1 Methodological conclusions

Spatial data and effects – The importance of using spatial techniques. In this thesis, exploratory analyses of spatial data turned out to be useful in investigating the spatial patterns in the proportion of commuting by bicycle per Belgian municipality (Chapters 2–3) as well as in the location of cycling accidents along the Brussels' network (Chapter 4). This notably allowed providing first insight into the factors that might play a role in explaining the observed spatial patterns (e.g. topography, availability/quality of cycle facilities, etc.). More importantly, they also helped in identifying the presence of global and local patterns of spatial autocorrelation and spatial heterogeneity (e.g. spatial outliers or clusters). For instance, in Chapters 2 and 3, a clear-cut north/south division of the Belgian municipalities was highlighted with respect to the proportion of commuting by bicycle. Also, analyses performed in Chapter 4 (i.e. network kernel density estimations and network K-functions) indicated that both reported and unreported cycling accidents spatially cluster along the Brussels' road network. Such results clearly suggested the presence of spatial autocorrelation and/or heterogeneity in the data. This is also confirmed by statistical tests performed in Chapters 3 and 5 (e.g. Moran's I, Lagrange Multiplier diagnostics, spatial Breusch-Pagan tests, etc.). It is well-known in the literature that, in the presence of such spatial effects, wrong statistical inferences can be obtained (e.g. biased estimates, misleading measures of fit, invalid tests, etc.). Special attention was then paid to account for spatial autocorrelation and heterogeneity throughout each of our modelling steps. Spatial modelling techniques were hence used in this thesis to correct for the presence of such effects, and our findings highlighted the importance of doing so.

In Chapter 3, spatial lag models turned out to be quite powerful in eliminating spatial autocorrelation and provided better fit than ordinary least squares (OLS) regressions. They also proved to be a better way of modelling than spatial error models, which was indicative of the fact that unmeasured/omitted explanatory variables were not at the root of spatial autocorrelation and, then, that our data collection was ample for our needs. The presence of spatial heterogeneity in the data – detected using the spatial version of the Chow test – was also corrected using White's correction and a disaggregated modelling strategy for the northern (Flanders) and southern parts of Belgium (Wallonia and Brussels), jointly with the spatial lag specification. As a result, the final model – referred here to as a spatial lag specification with regimes – provided a considerably better fit than the spatial lag and OLS models; the log-likelihood indeed increased from –102 (OLS) to 94 (spatial lag with regimes). The significance and magnitude of all the parameter estimates also greatly differed compared with OLS and illustrated how biased the estimates are when both spatial autocorrelation and heterogeneity are ignored. Interestingly, our results also showed substantial differences in the size of these estimates between the northern and southern parts of the country, which indicates that the variables may exhibit varying effects from one region to another. Last but not least, the addition of a spatial autoregressive component in the final model was suggestive of the existence of spillover influences between one municipality and its close neighbourhood. In other words, a municipality surrounded by others with high levels of commuting by bicycle is more likely to show high rates of commuter cycling (and vice versa). This not only indicates that social support for cycling could stem from the neighbourhood, but also that a virtuous circle could result from such spillover influences (in the long-term).

In Chapter 5, various spatial specifications were also used with the purpose to capture the effect of spatial autocorrelation. These specifications were conducted within a Bayesian framework as it provides several advantages over the frequentist/traditional estimation. Of interest for this thesis is notably the fact that it allows dealing with nuisance/random parameters (i.e. unobserved correlated and/or uncorrelated heterogeneity) in complex models. Our findings obtained using such a Bayesian computational approach however showed that specifications incorporating random effects (e.g. ICAR model) did not succeed in converging and provided insignificant parameter estimates for both random effects. At the opposite, models including an 'autocovariate' component at the first stage of the Bayesian hierarchy (i.e. autologistic models) conducted to the best results. This might indicate either that appropriate risk factors were included in the autologistic model, or that the spatial weight matrices used in the autologistic specification were the best to capture the unexplained variance associated with the presence of spatial autocorrelation. Interestingly, in the latter case, distance-based relationships were assumed in the autologistic model to reflect the neighbourhood influences between cycling accidents, whereas models incorporating random effects were based on more simplistic relationships (binary spatial weight matrices). This hence suggests that the definition of neighbourhood relationships throughout the construction of spatial weight matrices matters (note that it was also observed in chapter 3, but to a lower extent).

Network point pattern analyses – Reported *versus* unreported cycling accidents. In Chapter 4, spatial point pattern techniques – jointly with statistical tests for independence – have shown to be useful in exploring and comparing the spatial patterns of cycling accidents officially registered by the police (and compiled by DGSEI) with those unreported by police but collected through an open-based online registration survey (SHAPES survey; see Aertsens et al., 2010; de Geus et al., submitted). Given that cycling accidents are constrained to occur on a road network, this thesis took advantage of using recent point pattern methods extended to a network space (which is actually a one-dimensional space embedded in a plane). This extension to networks notably avoided drawing wrong inferences from the results (due e.g. to the over-detection of clustered patterns). In particular, special attention was paid to network K-function and network cross-K function methods in Chapter 4. The former method enabled us to depict the spatial distribution of both reported and unreported accident data sets, while the second one was used to examine whether or not unreported and reported cycling accidents occur in the vicinity of each other, and whether or not they have (dis-)similar locational tendencies with respect to specific infrastructure factors or facilities (e.g. intersections, schools, cycle lanes, tram tracks, etc.). Besides confirming findings from statistical tests for independence (e.g. Chi-Square adjusted tests) and centrographic methods (e.g. standard deviational ellipses), our results for Brussels indicated that unreported and reported cycling accidents overall exhibit similar spatial patterns along the road network and both cluster around similar infrastructures/facilities (except in some particular locations, such as traffic-calming areas; see Section 6.1.2). This hence suggests that enhancing the registration of cycling accidents would not necessarily provide further insight in unmeasured *spatial* factors associated with the occurrence of cycling accidents (at least, in Brussels). This has strong implications with respect to the interpretation of the model results obtained in Chapter 5, as it suggests that the statistical bias caused by underreporting might be overall slight. Moreover, official accident databases – such as these collected by police – may also probably serve as a good basis for orienting *in a global way* policy decisions and infrastructure investments in Brussels (although a more complete registration of cycling accidents is required if *local* safety treatments are intended by planners and decision makers). Last but not least, our findings also highlight the *importance of selecting an appropriate spatial subarea* for conducting network-based point pattern analyses. It is indeed demonstrated in Chapter 4 that our results strongly vary depending on the chosen spatial subarea (spatial clustering of cycling accidents tends to be more likely for increasing spatial subareas). This hence suggests that that great caution is required when conducting such network-based analyses on only one spatial subarea. At best, *several spatial subareas* should be used to check the consistency of the results.

Accident risk modelling and case-control strategies – Towards new research directions? In Chapter 5, particular attention has been paid to the estimation of the accident risk for cyclists and to the (spatial) factors that significantly affect this risk on the Brussels' road network. The direction this chapter has taken is however different compared with this opted within the framework of longitudinal surveys and traditional accident models (i.e. accident-frequency models or accident-severity models). Drawing inspiration case-control studies used in the research into epidemiology and ecology, a new methodological approach was here proposed to make possible *accident risk* modelling. This required the construction of a binary dependent variable, by coupling geocoded cycling accidents (DGSEI data, registered by the police) to *control sites* (i.e. locations where there is no reported cycling accident). Such controls were actually sampled along the 'bikeable' segments of the road network *and* as a function of a background exposure variable representing the bicycle traffic (which is estimated from a gravity-based approach). Of note is also the fact that black spots of cycling accidents were excluded from the 'bikeable' network in order to preclude the sampling of controls from the close vicinity of bicycle accidents. Once created, the binary dependent variable was then spatially intersected (or crossed) with potential risk factors manually digitised into a GIS. The resulting database – combining a binary dependent variable with attached risk factors – finally allowed modelling the accident risks for cyclists through the use of logistic and conditional autoregressive specifications (conducted here within a Bayesian framework). As mentioned above, the autologistic model turned out to be the best.

Such a modelling approach, based on a case-control strategy, provides several methodological advantages over traditional accident models (for which a large number of statistical biases are commonly reported) and longitudinal surveys (for which the selection of controls raises a number of questions that cast doubts about the validity of the resulting parameter estimates). These advantages are as follows: (*i*) the estimation/modelling of accident risks is made possible and is carried out in a more rigorous way compared to longitudinal surveys (especially as regards the choice and the representativeness of the control sites); (*ii*) the use of individual/point data avoids the need for arbitrary aggregation of accidents over some definite space (Diggle, 1990) and hence makes the analysis immune to the 'ecological fallacy'; (*iii*) the addition of controls avoids – or at least reduces – the small sample size problem; (*iv*) given that spatial variables are the only risk factors used here and as there is no classification of the level of severity, the underreporting issue related to cycling accidents is expected to affect to a lower degree the quality of the results (especially if this underreporting is spatially homogeneous); (*v*) there is no cross-model correlation between the different levels of injury severity (or collision types) as we did not take these into account; and (*vi*) the sampling of control points only depends on the location of black spots and on the spatial distribution of cyclists in the area of interest; if this latter remains unchanged throughout the years (and/or follows the same increasing trend over space), the use of out-of-date data does not bias the sampling of controls (and hence the results) since the intensity of this sampling is proportional to the exposure variable.

More interestingly, mapping the predicted values of the accident risk for a specific road trajectory on the Brussels' road network turned out to be useful in highlighting the locations at high risk of accident for cyclists. Compared to traditional black spot methods, this modelling approach provided three important advantages. First, it exploited *all the available information* (i.e. from the entire accident data set) to compute a predicted risk of accident for every point along the network, whereas black spot methods only use a small part of this information (for a definite accident, it only used the information relative to this accident and to the close neighbourhood). As a corollary, such a modelling approach hence holds the potential to *pinpoint locations where cycling accidents might have been unreported*. In Chapter 4, some locations on the network were indeed (rightly) highlighted as 'risky' by our modelling approach, whereas they were considered as 'safe' in black spot methods because no cycling accident was *officially* registered here (which is quite doubtful and suggests that underreporting might have been occurred here). Second, contrary to our modelling approach, black spot methods do not take into account the building/dismantlement year of road infrastructures. This could be a serious limitation when focusing on a definite period of time, since the black spot method could depict a particular location as 'dangerous' whereas it could not be anymore the case after some infrastructure treatments. Third, black spot methods do not consider the traffic direction and may highlight both sides of a street as dangerous whereas most cycling accidents cluster on one side only. Compared to our modeling approach (which took into account the traffic direction of cyclists as well as the street side where the infrastructures were built), black spot methods hence fail to give accurate precisions about the dangerous sides of the street and may then lead to erroneous inferences and decisions about infrastructures.

Concluding remarks. To sum up, this thesis provides four major methodological innovations through: (*i*) the use of spatial models to account for the presence of spatial autocorrelation and spatial heterogeneity in the data (mode choice research); (*ii*) the use of spatial point pattern methods extended to network spaces to explore and compare the spatial patterns of reported and unreported cycling accidents; (*iii*) the use of several spatial subareas to evaluate the impact of varying sizes of study regions (or varying network lengths) on the results obtained through network (cross) K-function methods; (*iv*) the use of controls, sampled along a network space and from an exposure variable, to construct a binary dependent variable (accident, no accident) that is in turn used in a spatial Bayesian model to estimate the risk of being involved in a traffic accident when cycling along the Brussels' road network. The first methodological innovation highlighted how biased the regression results are when spatial effects are ignored, and then provided strong support to the use of methods accounting for such effects when spatial data are used (especially in studies carried out in mode choice research, where the attention devoted to these effects is still limited, if ever, existent). The second methodological innovation allowed getting further insight in the spatial patterns related to the underreporting of cycling accidents, compared to official accident databases (e.g. how/where do unreported cycling accidents tend to locate along the network compared to reported cycling accidents?). Interestingly, this provided in-depth knowledge about the locations (and infrastructures) that are the most commonly associated with the occurrence of unreported cycling accidents. The third methodological innovation, in turn, emphasized the importance of selecting an appropriate spatial subarea for conducting network-based point pattern analyses, and then suggested for the first time in the literature that great caution is required when focusing on only one spatial subarea. Last but not least, the fourth methodological innovation provided a rigorous framework to estimate the accident risk for cyclists and to identify the most significant factors (infrastructures) influencing this risk. It notably offered a better tool than black spot methods to identify locations where the accident risk is the highest for cyclists and where cycling accidents might have been unreported or might still occur. Assuming that the risk factors (such as road infrastructures) have not been modified, such a methodology may hence greatly contribute to reduce the toll accidents take on public health as it holds the potential to prevent future (bicycling) accidents and allows cyclists choosing the safest route for their cycling trips. From a methodological and societal point of view, this last point probably constitutes the best achievement of this thesis and is hoped to provide a new research direction for traffic accident studies…

6.1.2 Empirical conclusions

Our exploratory and multivariate spatial analyses led to a plethora of results with regard to cycle commuting (Belgian municipalities), underreporting of cycling accidents and accident risks for cyclists (Brussels-Capital Region). Special attention is here paid on the major empirical results obtained throughout this thesis. These focus on: (*i*) the 'safety in numbers' effect, (*ii*) the spatial determinants of cycle commuting, (*iii*) underreporting of cycling accidents and locational tendencies, and (*iv*) the spatial factors of accident risks for cyclists in Brussels.

'Safety in numbers' effect in Belgium (municipalities). Our findings in Chapters 2 and 3 of this thesis are in line with the current research suggesting that higher levels of cycling are associated with lower rates of severe and fatal cycling accidents (see e.g. Jacobsen, 2003; Pucher and Dijkstra, 2003 ; Elvik, 2009). This latter hypothesis was first visually confirmed in Chapter 2 by clustering Belgian municipalities according to the proportion of commuting to work which was by bicycle *and* the risk of being seriously injured or killed when cycling to these municipalities. The results of this classification exhibited a clear-cut north-south division, suggesting that there are strong spatial differences in cycle commuting and accident risks between the Belgian regions. In the northern part of the country (Flanders), our findings showed that the municipalities overall have high proportions of cycle commuting and low rates of severe/fatal cycling accidents. Cycling is indeed part of the Flemish lifestyle, which may be explained by a number of factors interacting within a 'virtuous circle' that subsequently make the environment more attractive and safer for cyclists (see below for further details on these factors). Cyclists are then generally expected and respected by motorists in Flanders, since these latter often cycle themselves. In contrast, opposite results were obtained in Wallonia, where the environment is generally quite unsafe and unattractive to (potential) cyclists. Low proportions of cycle commuting and high risks of accident are indeed exhibited by the classification, which hence confirms the overall perception of danger Walloon inhabitants have about cycling. Of interest was also the fact that Brussels stood apart from the two other Belgian regions. It indeed showed low proportions of commuting by bicycle and low risks of severe/fatal accident for cyclists. Interestingly, these results do not support the fears/perceptions of danger people have about cycling in Brussels, as the risk of being seriously or fatally injured in a cycling accident is quite low here (which is explained by the fact that cyclists commute in an urban environment, where the speed differential between slow and fast modes is lower, compared to rural environments). Whatever the region, a number of factors – or spatial determinants – explained we observed a clear-cut north-south division in the country. They were clearly identified within the framework of Chapter 3 and are summarised here below.

Spatial determinants of cycle commuting in Belgium (municipalities). Chapter 3 aimed at identifying the spatial determinants of cycle commuting at the level of the Belgian municipalities, with the intent to subsequently provide sound recommendations for planners and policy makers. The results of our empirical analyses suggested that demographic, socio-economic, environmental and policy-related variables all influence the proportion of commuting by bicycle. For some of these variables, substantial differences were however exhibited in the magnitude and significance of the parameter estimates between the Belgian regions (Flanders *versus* Wallonia/Brussels). Income, gender and air pollution are variables for which the impact was only significant in Flanders, whereas variables related to the state of health, qualification and traffic volume (municipal/local roads) turned out to be significant only at the level of the Walloon municipalities. Among the *socio-economic* and *demographic determinants*, our results indicated that low median income and/or high proportions of working men are both associated with high rates of cycling to work in Flemish municipalities. At the opposite, the presence of high proportions of highly-qualified commuters is generally associated with low rates of commuter cycling in the southern part of the country (more especially in Wallonia, where Principal Component Analyses (not reported here) showed that positive associations exist between highly-qualified people, high median income, high car availability, and large commuting distances at the level of municipalities). Finally, the model results showed that being more than 45 years old and/or having one or more young children (≤ 5 years old) in the household decrease the likelihood of commuter cycling, whatever the region. As regards the *environmental* and *policy-related determinants*, our empirical analyses first revealed that, whatever the region, municipalities that are well-equipped and characterised by short commuting distances have high proportions of commuting of commuter cycling. Such results confirm several exploratory analyses conducted in Chapter 2 and validate the assumption that mixed-use and densely built environments (which are generally well-equipped municipalities) generate short trip distances and then encourage cycling. Second, our findings in Chapter 3 also reveal that a large part of the inter-municipality variation in cycle commuting is related to environmental aspects such as the relief, quality of cycle facilities and cycling accidents. Traffic volume on municipal roads however did not show any significant impact in Flanders, whereas it strongly discourages cycling in Wallonia and Brussels. As regards the topography, our results indicate that hillier terrains – when present – significantly discourage commuting by bicycle in all Belgian regions. Moreover, the lack of high-quality cycle facilities is shown to deter commuter cycling, as there is often no alternative but to cycle on-road in this case. Our results also reveal that the accident risk is negatively linked to commuter cycling, but to a lesser extent in Flanders. The assumption is that the high-quality of cycle facilities in Flanders strongly reduces the fears and annoyance of cycling into a heavy motorized traffic, which then puts the accident risk at the forefront of the resident's fears (so probably explaining the high value of the estimate for accident risks and the non-significance of traffic volume). In Wallonia and Brussels, due to the lack of appropriate cycle infrastructures, it is assumed that the first barrier with which *potential* cyclists are faced is the heavy traffic volume, not the accidents themselves (which in turn probably explains why the impact of traffic volume is significant, and even higher than this obtained for accident risks).

Apart from the spillover/mass effect exerted from the neighbouring municipalities on the propensity to cycle (see Section 6.1.1), our findings in Chapter 3 were mostly in line with the mode choice research. Interestingly, they corroborate some of the hypotheses put forward in Chapter 2 about the impact of several spatial factors on cycle commuting (e.g. distances, cycle facilities, built-up environments, etc.). They also show within a multivariate framework that high proportions of commuter cycling are associated with low risks of cycling accidents, which validates to some extent the results obtained in Chapter 2 as well as the previous statements referring to the 'safety in numbers' effect. Last but not least, our results (residuals of the final model) provide a useful tool to pinpoint both the municipalities that 'over-perform' in terms of bicycle use and those where there is still potential to encourage commuter cycling.

Underreporting of cycling accidents and locational tendencies in Brussels. Among other results, Chapter 4 of this thesis provided further knowledge about the spatial patterns of cycling accidents unregistered by the police, but collected through an open-based online registration survey (SHAPES survey). This was achieved by investigating *where* underreporting of cycling accidents mostly occurred compared to cycling accidents reported by the police. Zooming in the Brussels-Capital Region, our empirical results revealed that both unreported and reported cycling accidents show similar spatial patterns on a road network (i.e. they cluster with respect to each other along this network) and similar locational tendencies with respect to specific road infrastructures (such as intersections, bus and tram stops, etc.). This hence suggests that unreported accidents occur at rather similar locations to those that are reported. Therefore, it seems that registering accidents unreported by the police does not necessarily provide further insight in the *spatial factors* associated with the occurrence of cycling accidents. Exceptions are however reported in Chapter 4. Compared with reported cycling accidents, our findings indicate that cycling accidents are more likely to be unregistered in areas where the differential between the speed of slow and fast modes is reduced. Traffic-calming zones and streets located in the vicinity of schools, hospitals, cultural centres and shopping centres are examples of such areas where the speed of motorised vehicles is reduced through the implementation of speed limits, pedestrian zones and/or various physical obstacles (e.g. speed humps). In these areas, cyclists are more likely to be the only user involved in the accident and/or to incur slight injuries (with/without material damages). They hence generally do not feel the need to call the police, which results in a higher rate of underreporting by police. To sum up, our results hence suggest that traffic-calming measures have the effect of reducing the degree of accident severity and – as a corollary – the registration rate among (slight) cycling accidents.

Spatial determinants of accident risks for cyclists in Brussels (risk factors). In Chapter 4, the presence of potential collinearity problems between the risk factors did not allow drawing reliable conclusions on the separate safety effects related to infrastructures. Such collinearity problems were however avoided in Chapter 5 of this thesis. Our results are in line with the current traffic accident research (e.g. with respect to the increased risk of cycle facilities at intersection) and even provide further knowledge about the factors that were previously unexplored in a rigorous way in the literature (e.g. contraflow cycling, tram tracks, etc.). Figure 6.1 summarizes these by highlighting the factors that significantly affect the risk of cycling accident in Brussels and – then – that require great care when designing (new) infrastructures. Let us describe each of these findings as follows:

(i) *Bridges without any cycle facility*: increased risk of cycling accident when present. Hypothetical explanation: when no dedicated cycle facility is built on a bridge, cyclists are more exposed to sudden changes in road width (e.g. narrow space), road conditions (e.g. bridges are more prone to ice development) and visibility (curving of the bridge);

(ii) *High complexity*: the risk of cycling accident increases with 'complexity' (in the sense of the Elvik's law of complexity). Hypothetical explanation: cyclists and other road users face with a large number of information at the same time and must handle many visual stimuli at locations with an increased complexity (e.g. intersections). Cyclist's (and driver's) reaction time is then lengthened and driving errors are likely to be more frequent at such 'complex' locations, which may explain the greater risks of cycling accident observed here;

(iii) *Tram tracks*: increased risk of cycling accident when present. Hypothetical explanations:

a. *On-road tram tracks*: cyclists may get stuck in on-road tram tracks, resulting in a loss of control of the bicycle, and then in a fall. On-road tram tracks also impose cyclists to ride on specific places on the road and then probably increase the exposure of cyclists to other risk factors (such as the opening of car doors, aggressive drivers that are blocked behind, etc.);

b. *Tram tracks at intersections*: like on-road tram tracks, cyclists may get stuck in the tracks when riding in parallel to these in the intersection. Jointly with tracks, the presence of trams (and attendant public transport stops and users, in some cases) may also add some degree of complexity to the intersection;

(iv) *Cycle facilities at intersections*: increased risk of cycling accident when present (with different magnitudes of risk depending on the type of intersection and cycle facility). Hypothetical explanations (by descending order of importance/risk):

a. *Suggested cycle lanes built at right-of-way intersections*: the increased risk is likely to be caused by the non-respect of the right-of-way (mainly by other road users: 59%; cyclists: 10%) and/or the discontinuous aspect of the suggested cycle lanes. In the latter case, the use of chevrons and/or bicycle logos may indeed make these facilities less visible/expectable by motorists, especially when they are highly spaced within the intersection;

b. *Bidirectional separated cycle lanes built at yield/stop intersections*: motorists may have an inappropriate visual search pattern (i.e. they look at one direction only) when they cross bidirectional lanes at yield/stop intersections, which increases the risk of accident for cyclists riding in the opposite direction of the (parallel) traffic. This risk is even higher if the physical segregation of the cycle lane from the road brings an ill-founded feeling of safety to cyclists (which may persist at intersections);

c. *Marked cycle lanes built in roundabouts (outer lane)*: accidents frequently occur here when motorists leave or enter into the roundabout while cutting in on the trajectory of the cyclist riding on the marked lane;

d. *Unidirectional separated cycle lanes built at yield/stop intersections*: accident mechanisms are expected to be quite similar to those prevailing for yield/stop intersections equipped with bi-directional separated lanes. The only difference is that two-way cycling is not permitted on unidirectional lanes, although cyclists sometimes do it (such facilities are frequently built on either side of multi-lane and divided roads, which often deters cyclists from crossing the road to be in the right way);

e. *Marked cycle lanes built in signalised intersections*: motorists turning to an adjacent road may cut in on the (straight) cyclist's trajectory on the marked facility, so leading to the accident. Also, accidents might be explained by the fact that marked lanes built in signalised intersections generally position cyclists in the blind spot of heavy/large motorised vehicles. Of note is that the risk of accident is here lower compared to the above mentioned designs (which might be partly due to e.g. the presence of advanced stop zones for cyclists);

(v) *Roundabouts*: increased risk of cycling accident when present. Hypothetical explanation: accidents occur when motorists leave/enter into the roundabout while cutting in on the trajectory of the cyclist (who is merged into the stream of motorized traffic);

(vi) *Shopping centres*: the risk of cycling accident increases when riding closer to shopping centres. Hypothetical explanation: the intense pedestrian and/or motorised activity observed in the close vicinity of shopping centres increases the number of potential conflicting partners and situations, and then leads to a higher risk of accident for cyclists.

(vii) *Garages/parking driveways*: increased risk of cycling accident when present within a 100m network distance. Hypothetical explanation: motorists leaving or entering into a garage/parking driveway may cut in on the trajectory of the cyclist, who may eventually be hidden by close visual impediments (e.g. trees, hedges, etc.);

(viii) *Parked vehicles next to separated cycle facilities*: increased risk of cycling accident when separated cycle lanes (both types) are built close to parked vehicles ($\leq$ 0.8m). Hypothetical explanation: cycling accidents may be caused by the opening of car doors and by the attendant pedestrian activity on the separated cycle facilities (generated by the parked vehicles);

(ix) *Contraflow cycling (outside intersections)*: reduced risk of cycling accident when present. Hypothetical explanation: motorists may tend to adopt here a 'risk compensation behaviour', i.e. they may behave in a more cautious way due to an increased perceived risk when driving in such streets. Of concern is however the fact that intersections with such streets may result in a conflicting traffic situation, as motorists may be surprised to lie in front of (exiting) cyclists when entering into these streets;

(x) *Volumes of van and truck traffic*: increased risk of cycling accident with increasing volumes of van and traffic. Hypothetical explanation: on the one hand, the road legibility is as much reduced as the traffic is denser (as there is a great amount of information to handle). This hence reduces the cognitive capacity of cyclists and other road users, and then the ability to avoid accidents. Furthermore, vans and trucks are more prone to blind

spot problems and may also hide cyclists and put these out of sight of other road users (which then increases the risk of accident for cyclists);

At places where such factors are present, special attention to cyclists is crucial in order to reduce the risk (and the number) of accidents associated with cycling trips. This is even truer if several of these risk factors are present. For instance, a bridge equipped with on-road tram tracks and characterized by high van and truck traffic volumes is expected to be quite 'risky' for cyclists (more than in the case where only one of these risk factors – e.g. the on-road tram tracks – is observed).

Concluding remarks. To sum up, our empirical analyses conducted at the scale of the Belgian municipalities (part II) and on Brussels (part III) mostly provided further insight in: (*i*) the relationship between the proportion of commuter cycling and the risk of being seriously injured or killed when commuting to work in Belgium; (*ii*) the spatial determinants associated with the proportion of commuter cycling to work at the level of the Belgian municipalities; (*iii*) the underreporting of cycling accidents in Brussels; (*iv*) the spatial determinants (or risk factors) associated with the risk of being involved in a road accident when cycling in Brussels. The first set of results (Chapter 2) showed that there were strong spatial differences in bicycle use and the risk of accident between the Belgian regions. This in turn highlighted the importance several spatial variables might have in explaining such patterns. Second, variables for which the influence on commuter cycling was significant were then identified in Chapter 3. Our empirical analyses conducted in this latter chapter showed that socio-economic, demographic, environmental and policy-related aspects played an important role in influencing commuter cycling. Third, Chapter 4 of this thesis investigated *where* underreporting of cycling accidents mostly occurred compared to reported cycling accidents. Our findings led to two main recommendations. On the one hand, official databases of accidents should be analysed with great caution, especially as regards study regions where the number/length of streets equipped with traffic-calming measures is high (e.g. in the vicinity of schools, 30km/h areas, pedestrian and residential areas, etc.). On the other hand, registration efforts should be concentrated on areas where such (traffic-calming) measures are taken if the intent is to complete the current accident databases. Last but not least, Chapter 5 identified the spatial factors associated with an increased/reduced risk of cycling accident in Brussels. Only infrastructure- and traffic-related factors were retained by our empirical analyses. Infrastructure and policy measures relative to these factors (and their combinations) are of utmost importance since they are expected to provide the best safety benefits for cyclists.

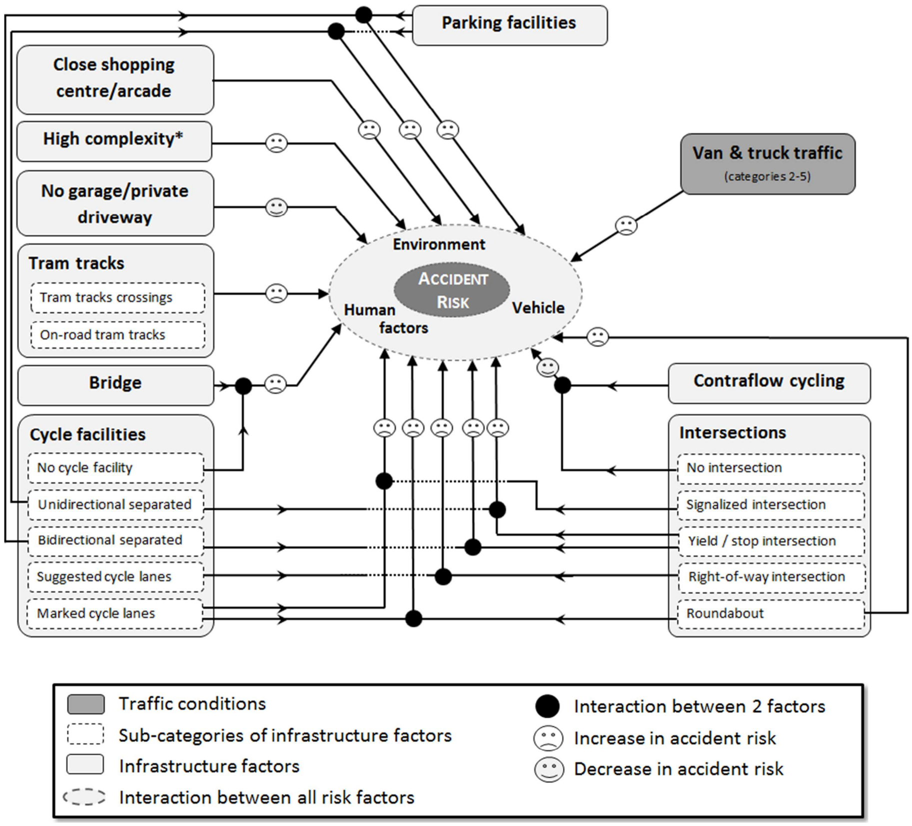


Figure 6.1: Significant factors (and their interactions) influencing the risk of cycling accident in Brussels. *Complexity is based on the Elvik's law of complexity (Elvik, 2006).

6.2 Policy implications and recommendations

Until relatively recently, transportation planners and policy makers are increasingly interested of obtaining science-based knowledge to encourage cycling and make it safer. As mentioned in Chapter 1, the use of the bicycle indeed holds the potential to take up the mobility, environmental and health challenges with which our society is faced nowadays. In line with such interests, the empirical analyses conducted in this thesis then provide several statistically-based recommendations that are helpful to support policies aiming at promoting more and safer cycling. Such recommendations are here categorized into five groups, corresponding to the well-known 5Es[1]. These are successively approached in the following subsections.

6.2.1 Engineering

Engineering can be very effective in increasing bicycle use and making it safer through better development, design and maintenance of cycle infrastructures, especially in areas where it is currently lacking. Our empirical analyses conducted at the scale of the Belgian municipalities (part II) and on Brussels (part III) clearly suggest that providing safe and well-designed road infrastructures (e.g. continuous cycle facilities) could prevent cyclists from falling or colliding with other means of transport. Our recommendations here distinguish general from specific engineering recommendations.

General engineering recommendations (part II). Our findings in Chapter 3 first suggest that, in Wallonia and Brussels, the provision of an extensive and high-quality cycling network would certainly reduce the numerous fears and safety concerns inhabitants have about cycling (Krizek et al., 2010). In particular, providing continuous, separate and well-maintained cycle paths could probably reduce the risk of accidents and mitigate the effects of traffic, as well as improving the general attitude commuters have towards cyclists (e.g. in terms of danger and societal status) (McClintock and Cleary, 1996). It could also reduce the exposure of cyclists to air pollution since even small 'separation distances' from the emission source significantly decrease the concentration of ultra fine particles (UFP) (Thai et al., 2008; Int Panis et al., 2010). Cycle networks should hence be planned so that the impact of deterrent variables (e.g. accident risks, slopes, traffic volume, air pollution) is reduced. Our results in Chapter 3 suggest that even small reductions in the daily mileage, the mean slope of the road network, or the risk of accidents could significantly increase bicycle use. This could be achieved by providing '*optimal paths*' for cyclists (i.e. alternatives to congested, sloping and/or hazardous roads). These paths could either be existing streets (e.g. quiet residential streets, without parking facilities) or new cycle lanes built along the road if high speed limits are permitted for motorists. Ideally, planners and engineers should design these latter so that they are separate from road traffic, but still allow cyclists and motorists to see each other, so that inexperienced and 'elderly' cyclists (who may behaviour inappropriately because of their age) are protected from motorised traffic but do not have an ill-

[1] The 5 Es are engineering, education, encouragement, enforcement and evaluation. The concept began in the 1970s in Odense (Denmark) and aimed at improving the safety of school children walking and bicycling to school (see Nielsen (1990) and PBIC (2007) for further information).

founded feeling of security. Our findings emphasizing the deterrent effect of high gradients also suggest that new cycle lanes should be made as flat as possible, or at least that any slopes should be long and gentle (so that the physical effort is reduced). For instance, new bridges specifically designed to enable cyclists to bypass dangerous or unpleasant situations should have gentle slopes. Including information about the topography on cycling maps and promoting the use of electric bicycles are other ways of 'bypassing' the negative impact of hilly terrain. Compared to car driving, electric bicycles – also called 'electrically-assisted-pedal-cycles' – not only yield a low-cost way to commute, but they also allow untrained individuals cycling in hilly municipalities.

The deterrent effect of high traffic volumes on cycling (reported in Chapter 3) could also be reduced by implementing strict parking policies and by regulating motorised traffic. Examples of such measures are parking and road capacity limitations (Pucher and Dijkstra, 2003; Pucher and Buehler, 2008). Land-use and urban design policies also hold the potential to reduce the dependence on car use and to create more economically efficient land use patterns. For instance, promoting dense and mixed-use development could reduce commuting distances and encourage cycling as well as other alternatives to car use. Redevelopment of urban areas (i.e. urban regeneration), promotion of bicycle storage facilities in blocks of flats (especially in dense residential areas), traffic-calming measures, and financial measures encouraging people to live in towns are some examples of such measures. Planning the urban centres and new housing centres in such a way that obstacles and cycling dangers are removed could also help increasing the safety and convenience of cycling (and then the use of the bicycle). Of importance is also the development of appropriate and secure bicycle facilities at the origins and destinations of the trip (e.g. cycle racks and secure lockers at transport stops), as it could increase users' satisfaction and encourage cycling and its integration with public transport (Martens, 2004, 2007; Pucher and Buehler, 2008). This could be particularly effective in large towns (such as Brussels or Antwerp) where vandalism and theft may deter cycling.

Finally, the promotion of folding bicycles and the implementation of a public bicycle sharing system would probably provide an efficient way to encourage cycling in urban areas, especially in Brussels and Antwerp where the potential for increasing bicycle use is still large (see Section 3.5.5.5) as inhabitants have generally little room to store their own bicycle (in the densest parts of towns, flats are smaller and few garages are available). Folding bicycles indeed allow carrying and storing the bicycle in small flats, while public bicycles can be hired in close stations. In the latter case, there is hence no need to store a bicycle in the flat, since it is parked in bicycle stations. Public bicycles also have the advantage not worrying about the maintenance of the bicycle. More importantly, both types of bicycles are also easily combined with other modes of transport, especially if public bicycle sharing stations are built near/in transports stops. In the long-run, the growing use of folding/public bicycles in urban areas could not only increase the use of non-motorised and public transport for commuting, but could also mitigate the deterrent effect of motorised traffic on cycling as well as the attendant negative impacts (air pollution, congestion, noise, etc.).

Specific engineering measures (part III). The focus is here put on safety-oriented recommendations aiming at making the use of the bicycle safer. Our results in Chapter 5 first suggest that, in Brussels, special attention should be paid to the bicyclist's safety when designing on-road tram tracks, bridges and/or 'major' intersections since these factors all increase the risk of cycling accident. It is all the more true if some of these risk factors are observed at a same location. In particular, major intersections are generally characterized by higher levels of complexity due to the presence of a dense crossing traffic as well as many road legs and signs. Whenever possible, they should be made more easily (and quickly) legible for all road users, e.g. by using the simplest possible signing or by decreasing the number of traffic lanes (and hence the intersection area). As regards tram tracks, crossable reserved tram lanes – or even physically segregated lanes – should be preferred to on-road tracks so far as possible. It could be profitable not only to cyclists (i.e. increased safety compared to on-road tracks) but also to public transport companies since such reserved infrastructures greatly improve the commercial speed of the vehicles (trams and buses). Bridges should also consistently be designed with a great care for cyclists in order to offset the increased accident risk caused by the reduced number and/or width of the road lanes. Building adjacent cycle facilities – separated with physical hurdles (e.g. barriers) – could probably reduce such a risk for the cyclists.

Cycle facilities should also be designed and built with great care, especially at intersections where the risk of having an accident is quite high for cyclists. In the case where investments devoted to the cycle facilities are limited, planners and decision makers should primarily give priority to the provision of high-quality infrastructure (i.e. continuous, visible and well-kept) rather than investing in an extensive network built in haste and carelessly. Separated cycle facilities should, for example, be designed in such a way that motorists get some time to see the cyclists before arriving at the intersection: while approaching it, the distance between the separated cycle facility and the adjacent road should be first reduced in order to favour a visual contact between the cyclist and the motorist, and then increased just some meters before the intersection (e.g. through a 2-5m deflection of the cycle facility from the main road) in order to give more time for both road users to see each other and to avoid the accident. As a complement, a sharp turning radius (90°) combined with a raised bicycle crossing and an advanced green light could also be implemented so that right-turning motorists are forced to slow down and cyclists get some advance over these latter to cross the intersection (Pucher and Buehler, 2008; Schepers et al., 2011). Concerning (on-road) marked and suggested cycle lanes, our results also suggest a quite high accident risk for cyclists at intersections equipped with such facilities. Making these more visible to motorists (e.g. using coloured pavements) is expected to reduce such a risk, especially for suggested lanes that are generally characterized by a discontinuous design. Also, small improvements at intersections may sometimes make all the difference in terms of accident risks. For instance, the installation of mirrors at signalized intersections may help lorry drivers to spot cyclists riding on cycle lanes and positioned in the blind spot of the vehicle, as well as they may remind them to check their own mirrors. Also, implementing advanced stop zones for cyclists here is expected to reduce the risk of accident associated with blind spot since they put cyclists into the view of motorists. Outside intersections, building (separated) cycle facilities in the 'door zone' of parked vehicles (< 0.8m) should be avoided as much as possible since the cyclists are exposed to a higher risk of accident due to the opening of car doors. A greater safety margin/distance (> 0.8m, or even > 1.2m) is here strongly supported in order to improve the bicyclist's safety. As regards the streets where contraflow cycling is permitted, the reduced accident risk reported here supports for a wider implementation of such a treatment in Brussels (and more generally, in most of the urban areas). Besides improving the safety, it has the advantage to require little investments and to be easily and quickly implemented in narrow streets (where there is no room to build cycle facilities). Great care should however be taken when designing these since the safety effect resulting from the treatment seems to be reduced at intersections. The use of (visible) marked cycle lanes or bicycle logos painted at the entrance of streets where contraflow cycling is permitted would probably be very useful in informing motorists that they could come face to face with cyclists, and hence in reducing the accident risk for cyclists here.

Last but not least, as illustrated in Figure 5.5 (Brussels), mapping the predicted values of accident risks along the *entire* road network would also allow cyclists choosing the safest route between an origin (e.g. residence place) and a destination (e.g. workplace, shop, school, etc.). Combined with other variables (such as the topography or the exposure to air pollution), optimal paths could then be determined for orienting cyclists to the safest and more comfortable routes. Providing such information to cyclists would also be of great interest for policy makers as it clearly holds the potential to reduce the health costs/risks associated with cycling. It is here thought that printed maps and applications dedicated to route planning (e.g. Google Map) could be efficient ways to disfuse such information to a large extent.

6.2.2 Education

Traffic education helps making road safety an integral part of the culture and lifestyle, as it is currently the case in the Netherlands, Germany and Denmark. Results in Chapters 2 and 5 showed that few motorists were respectful of cyclists in Wallonia and Brussels, which suggests that special attention should be paid to traffic education. In particular, our empirical analyses conducted in Chapter 5 clearly exhibit a higher risk of accident in several types of intersections equipped with cycle facilities. In particular, the non-compliance of traffic rules by motorists (e.g. non-respect of the cycle facilities) and – to a lesser extent – by cyclists (e.g. riding in the wrong way on a cycle facility) is found to be associated with the occurrence of cycling accidents at intersections (BRSI, 2006). This should be overcome by e.g. improving the driver training for motorists, teaching safe cycling practices, or by disseminating information aiming at improving the overall road safety (e.g. through safety campaigns) (Pucher and Dijkstra, 2003). Furthermore, bikepooling for the elderly or less-confident people as well as mobility education for local authorities and public services (e.g. administration, police) are some other measures that could improve both the road safety for cyclists and – as a corollary – bicycle use as a whole.

6.2.3 Enforcement

Enforcement strategies encourage all road users to adopt a more responsible driving style and to respect the rules of the road. In Chapter 5, our analyses conducted on Brussels suggested that accidents often resulted from the non-compliance of traffic rules, especially as regards the right-of-way and the speed limits in traffic-calming areas (which were here not found to reduce significantly the risk of accident for cyclists). Combined with traffic education, enforcement could make motorists more aware of and respectful towards cyclists. Collisions caused by drivers not respecting the right-of-way of cyclists (or triggered by the cyclists themselves) could be reduced by greater enforcement. As part of this strategy, more resources should be allocated in enforcement campaigns, and the punishment for violations of the traffic regulations should be far more severe, so that the perceived risk of being punished (following an illegal/dangerous manoeuvres) is increased. Furthermore, the implementation of bicycle patrols should be supported by decision makers in order to make police more mindful of the risks/deterrents with which cyclists are faced every day. Such patrols could then be very effective in improving bicyclist's safety, as well as in preventing bicycle thefts. Such patrols are especially required in Brussels and Wallonia, where aggressive driving and bicycle thefts are considered as important concerns by more than 70% and 40% of the inhabitants, respectively (Federal Police, 2006). Also, it is quite striking that, in Wallonia, more than half of all motorists were found to be going over the speed limit on 50 km/h roads, and nearly a quarter were over 60 km/h (2003–2006 period) (BRSI, 2008). Tackling such hazardous driving behaviour would obviously reduce the risk motorists constitute for cyclists.

6.2.4 Encouragement

Encouragement could also be useful to promote and increase cycling, especially in Wallonia and Brussels where the proportion of commuting by bicycle is quite low. Campaigns and mass events organised by public authorities and advocacy groups could be helpful in underscoring the health benefits as well as the improvements in the quality of life associated with bicycle use (reduction of noise and air pollution in the towns) (Pucher and Dijkstra, 2003). Decision-makers and health care professionals should also encourage individuals adopting a healthier lifestyle by integrating the use of the bicycle into the daily travel routines. Throughout our empirical analyses conducted in Chapter 3, we indeed observed that Walloon municipalities with low proportions of commuting by bicycle exhibited a high percentage of inhabitants estimating they had a bad state of health in 2001. Furthermore, public and private companies could also promote existing alternatives to the car, and try to make them competitive by providing financial incentives such as a mileage allowance or a company bicycle. Finally, taxes on fuel and automobile ownership/use also constitute some kind of encouragement to shift from car to alternative modes.

6.2.5 Evaluation

Evaluation allows for adjustments while a program of actions/measures is still in process and monitors if this provides the expected results and successfully responds to cyclists' needs. Evaluation can be conducted before, during and after the program by professional and neutral evaluators. It also means comparing the implemented cycling policies between different places (e.g. countries, municipalities or regions). This is actually what we did within the framework of this thesis. Although recommendations on eventual evaluation strategies do not follow from our empirical analyses, this thesis itself constitutes some kind of evaluation of the cycling policies and measures taken in Belgium and in the Brussels-Capital Region. It is here hoped that it will help planners and policy makers to evaluate the current bicyclist's situation and will then support adequate policies encouraging more and safer cycling in Belgium.

6.2.6 Concluding remarks

For transportation planners and policy makers, this section provides several strategies which may be useful in making bicycle use safer and in encouraging commuters to shift from car to bicycle. Such strategies may not only enhance the environmental quality, but they also hold the potential to improve the performance of the labour market and the local/regional economic development (e.g. through the establishment of new companies and residents attracted by the resulting quality of life). These strategies are, however, generally not efficient when implemented on their own. For instance, policies aiming at reducing the traffic volume in urbanized areas (e.g. urban toll) would have unexpected safety consequences for cyclists if they are done on their own (i.e. without traffic calming measures, traffic education, etc.), since the ability of vehicles to travel faster is increased. At worst, they may lead to adverse effects for the cyclists' safety and decrease bicycle use (Shefer and Rietveld, 1997; Noland and Quddus, 2004). Consequently, planners and policy makers should be aware that only a combination of several measures (enforcement campaigns, traffic education, improvement of cycle facilities, etc.) will really lead to an increase in cycle commuting (Pucher et al., 2010).

Also note that the recommendations provided in this section do not result in the same degree of achievement as some measures are more complex than others to implement, depending on the costs, administrative tasks, public acceptability, or policy objectives. For instance, the wider implementation of streets where contraflow cycling is permitted is far easier to achieve than land-use measures aiming at promoting a dense and mixed-use development of activites (which may involve high research costs, time-consuming administrative tasks, and voluntary policies). For each measure, Table 6.1 gives an evaluation of the degree of achievement, as well as it yields further information about the level at which it could be implemented (e.g. municipal, regional, network, etc.) and on potential target places where measures should be taken first and foremost.

Table 6.1: Policy recommendations

Category	Measures	Objective(s)	Thesis-related factor(s)	Achievability*	Level(s) of implementation	Target places?
Engineering						
	Provision of an extensive and high-quality **cycling network** (e.g. continuous, well-maintained, visible, etc.)	Mitigating the negative impacts associated with high objective / perceived accident risks, high motorised traffic volumes, and air pollution, and then encouraging bicycle use	Dissatisfaction of cycling facilities, accident risk, traffic volume (chapter 3)	Intermediate	Regional	Connections between specific locations (towns, facilities, etc.)
	Provision of '**optimal paths**' for cyclists	Providing alternative roads to congested, sloping and/or hazardous roads between specific origins and destinations	Accident risk, traffic volume, topography, contraflow cycling, etc. (chapters 3 & 5)	High	Regional, municipal	Workplaces, schools, transport stops
	Provision of **secure bicycle facilities** (racks, changing facilities, etc.) at the origins and destinations of the trip	Increasing user's satisfaction, encouraging commuter cycling and its integration with public transport	Bicycle theft, urban hierarchy (chapter 2)	High	Regional, agglomeration, municipal	Workplaces, schools, transport stops (especially in large towns)
	Implementing **traffic-calming measures** and / or traffic **restrictions** (in target places only)	Regulating motorised traffic and reducing the differential speed between slow and fast modes of transport	Traffic volume, accident risk (chapters 3 & 5)	Intermediate	Regional, agglomeration, municipal	Residential wards, schools, hospitals
	Favouring the implementation of **public bicycle sharing systems**	Creating a supportive environment for cycling (e.g. through a mass effect)	Spatially lagged variable (chapter 3)	Intermediate	Agglomeration	Large and medium-sized towns

continued on next page

continued

Category	Measures	Objective(s)	Thesis-related factor(s)	Achievability*	Level(s) of implementation	Target places?
	Designing legible **intersections** for all road users (e.g. simplest possible signing, reduced number of traffic lanes, etc.)	Improving the legibility of (complex) intersections and reducing the accident risk	Complexity, intersection-related factors (chapter 5)	Intermediate	Local (street network)	Major intersections, with a dense crossing traffic and a large number of road legs
	Preferring **crossable reserved tram lanes** to on-road tracks	Reducing the (high) accident risk associated with the presence of on-road tram tracks. Note that this measure also has the advantage to increase the commercial speed of the public transport vehicles	On-road tram tracks, crossable reserved tram lanes (Chapter 5)	High	Local (street network)	On-road tram tracks
	Designing **bridges** in such a way that special attention is devoted to cyclists (e.g. by building adjacent cycle facilities)	Reducing the (high) accident risk associated with the presence of bridges (unequipped with cycle facilities)	Bridge & no cycle facility (chapter 5)	High	Local (street network)	Bridges without any cycle facility, with a reduced number and/or width of the road lanes
	Designing **cycle facilities** with great care, especially **at intersections** (e.g. through the implementation of raised bicycle crossings, coloured pavements, mirrors at signalized intersections, advanced stop zones, etc.)	Reducing the (high) accident risk associated with (specific types of) intersections when cycling on (specific types of) cycle facilities	Intersection-related factors & Cycle facility-related factors (chapter 5)	Intermediate	Local (street network)	Cycle facilities at intersections

continued on next page

continued

Category	Measures	Objective(s)	Thesis-related factor(s)	Achievability*	Level(s) of implementation	Target places?
	Building separated cycle facilities outside the '**door zone**' of parked vehicles (< 0.8m)	Reducing the (high) accident risk associated with the opening of car doors when cycling on separated cycle facilities	Proximity parking-cycle facility (chapter 5)	High	Local (street network)	Cycle facilities built in the 'door zone'
	Supporting for a wider implementation of streets where **contraflow cycling** is permitted (with however great care at intersections)	Making the use of the bicycle more convenient (by reducing the travel time) and safer. Great care should however be taken at intersections with streets where contraflow cycling is permitted	Contraflow cycling (chapter 5)	Very high	Local (street network)	One-way streets in agglomerations
	Promoting **dense and mixed-use development**, and favouring the **redevelopment of urban areas** (i.e. urban regeneration)	Reducing commuting distances and encouraging active modes of transport (e.g. cycling and walking)	Commuting distance, urban hierarchy (chapters 2 & 3)	Very low	Regional, agglomeration	Large and medium-sized towns
	Promotion/provision of **bicycle storage facilities** in blocks of flats	Encouraging the use of active modes of transport	Town size (chapter 2)	Very high	Regional, agglomeration	Dense residential districts in large / medium-sized towns
Education	Improving the **driver training** for motorists	Making road safety an integral part of the culture and lifestyle	Accident risk (chapter 2), intersection-related factors & Cycle facility-related factors (chapter 5)	High	Regional, agglomeration, municipal	Workplaces, local authorities, public services, schools, driver trainings, mass events
	Teaching **safe cycling practices**			High		
	Disseminating information through **safety campaigns**			High		
	Bikepooling for the elderly and / or less confident people			Very high		

continued on next page

continued

Category	Measures	Objective(s)	Thesis-related factor(s)	Achievability*	Level(s) of implementation	Target places?
Enforcement	Allocating more resources in **enforcement campaigns**	Increasing the perceived risk of being punished (following an illegal/dangerous manœuvre)	Accident risk (chapter 2), intersection-related factors & Cycle facility-related factors (chapter 5)	Intermediate	Regional, police zone	Residential districts, schools, intersections with cycle facilities
	Making the **punishment for violations** of the traffic regulations far more severe (especially for some of these)	Increasing the perceived risk of being punished (following an illegal/dangerous manœuvre or behaviour)	Accident risk (chapter 2), intersection-related factors & Cycle facility-related factors (chapter 5)	High	Regional, police zone	Residential districts, schools, intersections with cycle facilities
	Implementing **bicycle patrols**	Making the police more mindful of the risks/deterrents with which cyclists are faced everyday	Accident risk (chapter 2), intersection-related factors & Cycle facility-related factors (chapter 5)	Very high	Regional, police zone	Residential districts, schools, intersections with cycle facilities
Encouragement	Organizing **campaigns and mass events** aiming at promoting bicycle use	Emphasizing the health benefits and the improvements in the quality of life associated with cycling, in order to encourage it	Bad health (chapter 3)	High	Regional, agglomeration, municipal	Large and medium-sized towns, workplaces, health care professionals
	Promoting existing alternatives to the car and making them competitive by providing **financial incentives** (e.g. mileage allowance, company bicycle, taxes on fuel and automobile ownership, etc.)	Encouraging and rewarding a modal shift from car to alternative modes	Traffic volume (chapters 3 & 5)	High	Regional	Workplaces, schools

continued on next page

continued

Category	Measures	Objective(s)	Thesis-related factor(s)	Achievability*	Level(s) of implementation	Target places?
Evaluation	Monitoring if a program of actions/measures provides the expected results	Evaluating how a program of actions/measures successfully responds to cyclists' needs, and undertaking adjustments (in these actions/measures) if necessary	-	Intermediate	Transnational, national, regional, municipal	Workplaces, schools, municipalities, agglomerations, regions, countries (mobility plans, transport and land-use policies)
	Comparing the implemented cycling policies between different places					
Others	Provision of information about the **topography** (e.g. through cycling maps)	Reduce the physical effort associated with cycling	Slopes (chapter 3)	High	Regional, municipal	Schools, workplaces, hilly municipalities
	Promoting the use of **electric bicycles** through financial incentives	Reduce the physical effort associated with cycling	Slopes (chapter 3)	High	Regional, municipal	Hilly municipalities
	Implementing strict **parking policies** (parking and road capacity limitations)	Regulating motorised traffic in agglomerations	Traffic volume (chapters 3 & 5)	Intermediate	Regional, agglomeration	Large and medium-sized towns

- : measure not derived from our results

* To be interpreted with great caution since the degree of achievability of a measure depends on a large range of factors. This degree is here evaluated on the basis of the author's knowledge as regards 4 factors: cost of the measure (expected building costs, maintenance costs, workforce costs, etc.), study/research requirements (expected time budget required to undertake research, studies, etc.), administrative tasks (expected administrative and political difficulties, e.g. as regards the period of implementation or the administrative tasks required to achieve the measure), and degree of acceptability (expected popularity among all road users, e.g. car drivers, cyclists, or public transport users).

6.3 Limitations of this thesis

This thesis is not without weakness. Additionally to some previously raised limitations in part III of this thesis, it is worth to mention that other major technical and methodological issues were experienced throughout this thesis and merit further research. The focus is here put on data (Section 6.3.1) as well as on methodological and technical limitations (Section 6.3.2) encountered throughout this thesis.

6.3.1 Data limitations

Although a wide range of data were collected throughout this thesis, the data collection is still far from being exhaustive. Some of the data specifically related to cycling and accident risks for cyclists were indeed not collected, mostly because of confidentiality reasons, unreliable information, and/or time constraints. For instance, weather- and/or climatic-related variables were not used in part II of this thesis since the data were collected over a limited number of measurement stations. This resulted in a spatially poor representativeness of the data and precluded us from using these within the framework of our empirical analyses (which are conducted at the scale of the Belgian municipalities). In the third part of this thesis, some factors were also deliberately ignored because they were affected by frequent infrastructure changes during the period under study (e.g. advanced stop zones for cyclists), and/or because it required time-consuming field observations to obtain reliable data (e.g. traffic lights for cyclists). Some of the data manually digitised into our GIS also raise some questions about their validity. Although the digitization process was carried out over several years and drew information from several data sources (e.g. cycling maps, BRIC, etc.), it does not claim to be as precise as field observations. Examples of data being particularly concerned by these issues are parking areas since their delineation clearly depends on the temporal variation in the parking behaviours and, then, on the moment at which the orthophoto has been taken. Parking occupancy indeed strongly varies according to the day (e.g. weekdays *versus* week-end) and hour of the day (e.g. off-peak *versus* peak hours). This hence forced us to make sometimes strong assumptions about the actual parking occupancy. For several infrastructure-related data, there was also seldom, if ever, information on the implementation/dismantlement year (which was only assessable within several months). Despite the fact we kept watch over eventual infrastructure changes, there is hence some likelihood that encoding errors might have occurred and might have affected the results obtained in Chapters 4 and 5.

Of particular attention is also the fact that most of the demographic, socio-economic and mobility-related data used in part II of this thesis come from the Belgian Census of Population, conducted by the DGSEI in 2001 (i.e. about 10 years ago). Some of the data extracted from this latter can then be considered as quite obsolete for our purposes. As it was compiled for the last time by the DGSEI (surveys now replace the census), it however still constitutes the most recent database covering the *entire* Belgian population and, then, providing the finest spatial representativeness. Given that the focus is here put on the spatial analysis of data, it was then decided to rely on this census despite its relative obsolescence.

Several issues may finally be raised as regards the traffic accident database we used throughout this thesis. First, in Chapter 2, the absence of information about the trip purpose is likely to bias our results (over-estimation of the severe/fatal accident risks). It is here advised that, in a near future, further information about this variable (trip purpose) should be collected when registering cycling accidents. Second, the underreporting of cycling accidents is expected to affect our results obtained in Chapter 5, although to a lesser extent compared to other statistical methods. Chapter 4 indeed suggested that unreported and reported cycling accidents exhibit similar locational tendencies with respect to specific infrastructures and facilities. Third, insufficient and/or imprecise information may be associated with both reported and unreported accident databases, which may subsequently affect our empirical analyses as they are conducted at the network level (and then need detailed information on the accident location and mechanisms). Doubtful information was however eliminated as far as possible from the accident databases, thus mitigating the risk to make wrong inferences.

6.3.2 Methodological and technical issues

Some methodological gaps were noted within the framework of this thesis. First of all, in part II, the choice of Belgian municipalities as basic spatial units raise some questions about their relevance in reflecting homogeneous environments (with regard to e.g. the human activities, the natural environment, the socio-economic characteristics, etc.). However, such a choice was constrained by the level at which data on explanatory variables are available. The lack of high-resolution information for some of our variables (as regards e.g. traffic volume, cycle facilities, air quality, etc.) indeed required aggregating the data at the level of municipalities. Given that the results of empirical analyses may vary as a function of the size of spatial units (see Chapter 1, MAUP), it would then be of particular interest to undertake spatial analyses at different levels of aggregation (especially at finer levels, in the case where high-resolution data are available). Such a multilevel analysis would in turn evaluate the effects of different levels of aggregation on our results (thus confirming or invalidating these latter).

Secondly, in part III of this thesis, there is also some inconvenience to delineate our study area on the basis of the administrative boundaries of the Brussels-Capital Region. Focussing on regional boundaries indeed implies that our analyses are performed in a ‘closed system’. They hence assume that there is no neighbourhood (and, then, no external influence). Such an assumption is not realistic as it ignores the potential effect of factors having an influence extending beyond administrative boundaries (e.g. shopping centres in part III). These ‘edge effects’ indirectly results from the regional structure of Belgium. The availability and the definition of data may indeed differ from one region to another, which either precludes performing analyses outside regional boundaries or imposes concentrating a greater amount of time on the data collection (especially if the intent is to work at the level of the Brussels’ urban agglomeration, which includes municipalities embedded in the three Belgian regions). From a methodological point of view, it is expected here that such edge effects may lead to an underestimation of the impact of some ‘peripheral’ factors (i.e. those located in the periphery of the study area). Moreover, they are likely to hamper the ability of cross-K function methods to detect a significant clustering (or dispersion) of cycling accidents around definite factors (especially those observed in the periphery of the study area).

Thirdly, in Chapter 4, the inability to account for the street side and building/dismantlement year of infrastructures, as well as the computational intensiveness related to the K-function and cross-K function analyses (especially with large datasets and/or high network lengths) are limitations that cannot be solved in a straightforward way in SANET. Although many improvements have been recently achieved with regard to spatial network analyses of point patterns, there is indeed still no research in the literature proposing to account for high levels of details on the street side where the accident actually occurred or on the temporal evolution of road infrastructures (e.g. implementation or dismantlement of infrastructures). This may clearly bias our network kernel density estimations (as they aggregate the estimation for both street sides), as well as our results obtained using the network cross-K function method (because they use infrastructure-related data for which modifications may occur). Focussing on more technical aspects, it was noted in SANET that the computational intensiveness strongly depends on the length of the network, the number of basic (e.g. road infrastructures) and non-basic points (i.e. cycling accidents), and – to a lesser extent – the computer specifications. Overall, short to moderate computation times were required at the scale of the Pentagon's street network (i.e. about 1-100 minutes depending on the number of points), whereas the First and the Second Crowns led to moderate to (very) high computation times (i.e. some minutes to about 6 days). For dense networks and large point datasets, the high computational intensiveness of SANET then clearly limits the number of spatial network analyses that can be conducted during the research period.

Fourthly, the use of individual data in Chapter 5 also has some major drawbacks. Depending on the requirements about the quality of the data (i.e. road network, local risk factors, etc.) and the size of the studied area, the data collection may be time-consuming since it requires collecting additional data for the controls (or for the whole studied area). Moreover, the quality of the results is strongly constrained by the method of selection of the controls as well as by the formulation of the potential index (e.g. choice of the impedance function). Although the potential index for bicycle traffic (i.e. the exposure variable) is shown to be quite representative of the observed cycling trips in Brussels, it may still be improved by assigning a preferential direction of travel into its specification (e.g. towards the town centre) and/or by considering cycling trips carried out for purposes other than commuting (e.g. leisure, shopping). Also, the validity of the results has not been tested for different types of sampling methods (e.g. regular sampling *versus* stratified sampling) and for various ratios of controls to cases (i.e. for a varying number of controls M_0 against the number of cases n, e.g. M_0/n, $2.M_0/n$, $10.M_0/n$, $100.M_0/n$, etc.). Despite the fact that the sampling of controls is based on well-founded theoretical bases and performed on a thoroughly constructed exposure variable, it would merit further investigation to implement such an analysis of sensibility for different control data sets.

Finally, it is worth of note that the bulk of this thesis is limited to the spatial aspects of the data. Several factors related to the individual attributes (e.g. preferences, attitudes, etc.) were not analysed here. For instance, in Chapter 5, the use of control sites indeed imposed us to put human- and vehicle-related factors aside given that the random assignment of these factors to control sites was considered as rather tricky (in the sense it is expected to bias the model results). Also, in Chapter 3, our data are spatially aggregated and then ignore some important individual components that could play a role in explaining the use of the bicycle for commuting to work (e.g. work schedule, dress code, etc.). Such 'aspatial' aspects would therefore merit further research in the future.

6.4 Perspectives for future research

While this thesis addresses several empirical and methodological issues, it raises new research questions and delivers new directions for traffic accident research. First, we suggest that further knowledge should be accumulated as regards the spatial effects of commuter cycling (i.e. spatial autocorrelation and heterogeneity). While such effects are observed at the scale of the Belgian municipalities, nothing is known about their potential existence at other levels of aggregation and/or across other study areas (e.g. countries, regions, etc.). This would not only give more clues about the range of scales at which diffusion processes occur, but it also holds the potential to provide further insight in the factors that determine such processes. Focussing on finer scales, for instance, would help to determine whether or not there is some kind of neighbourhood effect between the residents of adjacent districts (e.g. social support, influence of neighbouring pro-cycling policies, etc.).

Second, Chapter 2 highlighted the fact that high levels of commuter cycling are associated with low risks of becoming seriously injured or killed when cycling. However, nothing or little is known about the relation of cause and effect underlying such an association. Complex inter-relationships indeed exist between the different underlying factors (e.g. culture, visibility of cyclists, investments in cycling facilities, actual and perceived risk of accident, etc.) and it would merit further research to investigate the feedback effects that lie at the root of such a ‘safety in numbers’ effect. This would notably allow confirming (or not) the fact that high levels of cycling result in lower accident risks for cyclists. If such a relationship is confirmed, factors having either direct or indirect effects would also be identified, and then further knowledge would be available for planners and policy makers to initiate and/or maintain a virtuous circle.

Third, our empirical analyses conducted in Chapter 3 focus on Belgian municipalities and do not examine spatial factors influencing commuter cycling at different levels of aggregation. In particular, at finer scales of analysis, it is expected that different spatial factors would play a role and that ‘well-targeted’ policy recommendations would be established to encourage the use of the bicycle. Special attention should also be paid to the relative importance between individual characteristics and ‘trip-related’ data (e.g. infrastructures between the origin and the destination of the trip). Conducting statistical analyses at the individual level (using e.g. logistic models) would indeed be helpful in providing further knowledge on how specific types of factors (e.g. cycling infrastructures) would influence the choice of the bicycle as mode of transport, relatively to individual characteristics. It would however require a time-consuming data collection since ‘trip-related’ data are seldom available. Although it still provokes some debate in the literature, multilevel modelling could also be of interest as it allows incorporating several hierarchical levels of analysis in the model, with the aim to separate compositional effects from contextual ones (see e.g. Vanoutrive et al., 2010). When modelling the modal choice of the bicycle in a given spatial unit (e.g. municipality), such a multilevel modelling could then separate the effect of individual characteristics (e.g. culture, income or age of the individuals) from the neighbourhood/environmental effects (such as the risk of having a cycling accident in a given municipality).

Fourth, our data used within the framework of Chapter 3 do not account for: (*i*) the combination of the bicycle with other transport modes in the dependent variable (multimodality), (*ii*) the presence of public bicycle sharing systems (as independent variable). On the one hand, the selection of cyclists-only (i.e. cyclists who used the bicycle as only mode of transport) is motivated by the fact that, in the 2001 census, commuting distances are reported for the entire journey without any distinction of the transport mode (i.e. only the total distance is reported). On the basis of exploratory spatial data analyses (not reported here), it turns out that accounting for such a combination (bicycle-other mode) would have increased the shares of cycle commuting in some municipalities, especially those where high-quality public transport is present. Several towns equipped with major railway stations (e.g. Gent, Kortrijk, Etterbeek, Ixelles, Ottignies-Louvain-la-Neuve) indeed exhibit higher shares of cycle commuting when attention is paid to multimodality. Hence, accounting for proxy variables related to the urban environment and/or to the accessibility to railway stations would probably have been useful in explaining the variance associated with multimodal trips (bicycle-other mode). Within the framework of this thesis, it is however expected that accounting for such multimodality in commuting trips would not have strongly affected our results[49]. On the other hand, the fact that public bicycle sharing systems are not taken into account in Chapter 3 does not bias our results as such systems were not implemented before 2006 in Belgium, while our data belong to the period 2000-2005. Our results are hence valid for this latter period. However, Brussels (2006) and Antwerp (2011) were recently equipped with these systems, and Namur will be the next Belgian town to benefit from these (during spring 2012). It is hence questioned here whether or not our results would be still valid after the implementation of such systems. Although we do not know of any statistics about the impact of these systems on the modal share of cycling in Belgium, studies conducted in foreign countries suggest that cities equipped with such systems experienced – immediately after

[49] However, the use of more recent data would probably have changed the results as public bicycles and folding bicycles both have a growing success in urban areas and allow combining the bicycle with other modes of transport.

the launch and later – a considerable increase in bicycle use (e.g. +80% cycling trips in Lyon, from the launch in June 2005 to May 2009; +70% cycling trips in Paris from the launch in July 2007 to June 2008) (City of Paris, 2008; Greater Lyon, 2009). Hence, it would be of interest to analyse the impact public bicycle sharing systems have on bicycle use (e.g. what is their relative importance/impact compared to other factors?). It is here expected that the initial shares of cycling and contextual factors (e.g. dense residential ward, in which bicycle storage facilities are lacking) would strongly influence the potential use of such public bicycle sharing systems. In municipalities where initial shares of cycling are low and where bicycle storage locations are lacking, it is expected that predicted values of the model in Chapter 3 would be affected if public bicycle sharing systems are introduced in a municipality[50]. In such a case, collecting variables summarizing the presence/absence/accessibility of these systems (e.g. ratio between the number of public bicycles and the number of inhabitants, per municipality) would probably be useful to enhance the model fit.

Fifth, in the case where accurate data are available at the local level (with regard to road infrastructures and bicycle traffic flows or cyclist's living places), the case-control approach implemented in Chapter 5 for Brussels is expected to be easily transposable to other areas. In particular, it would be interesting to analyse how the accident risks (along the network) vary from one study region to another, especially if the design and the availability of infrastructures differ (e.g. in terms of cycle facilities, presence/absence of tram tracks, etc.). Our empirical analyses conducted in Chapter 5 indeed do not have the pretention to provide a generalizable answer to the safety effects related to each risk factor. In Brussels, high-quality cycle facilities are lacking and the popularity of cycling is still low (about 4%) compared to other European towns sharing the same socio-economic, demographic, environmental and mobility characteristics. It is clearly not representative of environments where the use of the bicycle and the investments in cycle infrastructures are far higher (as it is the case in e.g. Flemish and Dutch towns). Similarly, rural environments are expected to lead to different results as they are not characterised by the same road infrastructures (e.g. absence of tram tracks and low traffic volumes). Whatever the final choice on the study region, we think that longitudinal surveys paying special attention to the spatial dimension of the sampling design could be helpful in estimating such risks in a reliable way. Such a spatial dimension clearly matters and should not be ignored if the intent is to provide sound recommendations to planners and policy-makers.

[50] Also note that similar results would probably be observed as regards folding bicycles.

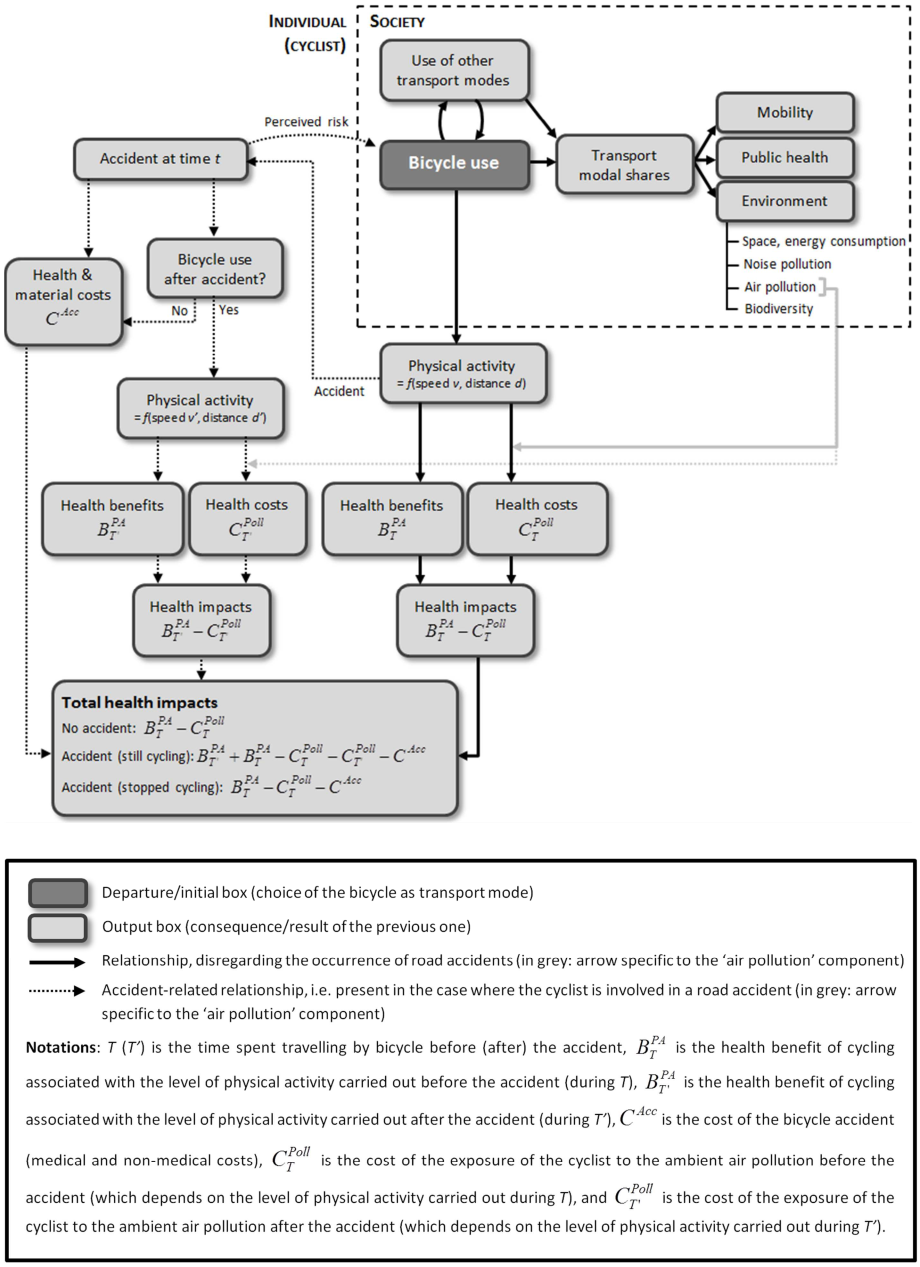


Figure 6.1: Total health impacts of bicycle use – Under-estimation of the health costs of cycling accidents

Last but not least, the high underreporting rate of minor/slight injuries observed in Brussels (Chapter 4) suggests that current estimates of health costs/benefits associated with cycling are probably biased due to the underestimation of the total cost of bicycle accidents. Such an underestimation is even expected to be higher as most heath impact assessment studies do not account at all for such minor/slight injuries; they mainly estimate the health costs/benefits from bicyclists' fatalities (see e.g. Woodcock et al., 2009; de Hartog et al., 2010; Rojas-Rueda et al., 2011). Yet, these latter – although highly (negatively) valuated in monetary terms – form the 'tip of the iceberg'. On the contrary, minor/slight injuries *officially* account for more than 95% of the bicycling accidents in Belgium (which could amount to more than 99% since minor/slight injuries are strongly underreported) (see Chapter 4). Despite the fact it was recently demonstrated that the health benefits of cycling exceed the health risks when considering both fatal and minor/slight injuries (Rabl and de Nazelle, 2012), health impact assessment studies still disregard some of the consequences injuries may have on the level of physical activity of the (injured) cyclist after the accident. They indeed make the strong assumption that the cyclist (i.e. the individual itself) does not modify/adapt his/her level of physical activity after the accident. However, cycling accidents may have important repercussions on the 'future' level of physical activity of the (injured) cyclists, depending on e.g. the injury severity, the circumstances of the accident, or the psychological consequences. For instance, a physical invalidity may result from injuries and may preclude the cyclist/individual from cycling during a given period of time (following the accident), or even during the entire life in the case where physical invalidity is permanent. Also, having an accident during the night-time or during a windy or foggy day may encourage the cyclist adapting its travelling behaviour by choosing cycling during daylight or 'normal' weather conditions. A high variability in the level of physical activity may hence result from a cycling accident, which then questions the validity of the estimates obtained through current health impact assessment studies as the total cost of bicycle accidents is expected to be underestimated in such a case.

Figure 6.1 illustrates well this issue. Let us consider that an individual switched to bicycling and had an accident at time *t*, after having devoted travelling a total time *T* by bicycle. In the case where the individual stops cycling or strongly reduces his/her level of physical activity because of a cycling accident (due to e.g. invalidity, psychological consequences, influence of the family, etc.), the total health impacts of cycling might be negative since the health benefits of physical activity accumulated during *T* and *T'* (B_T^{PA} and $B_{T'}^{PA}$, respectively) might not exceed the total costs of the bicycle accident (C^{Acc}) and exposure to air pollution (C_T^{Poll} and $C_{T'}^{Poll}$). This is more likely to be true if the individual stops cycling because of the accident, as the total health impacts for the cyclist are positive only if $B_T^{PA} > C_T^{Poll} - C^{Acc}$ (which is likely to be negative if high costs resulted from the bicycle accident). If the individual reduces his/her level of physical activity after the accident, the total health impacts are positive in the case where $B_{T'}^{PA} + B_T^{PA} > C_T^{Poll} - C_{T'}^{Poll} - C^{Acc}$. Finally, if the individual does not incur any cycling accident, the total health impacts ($= B_T^{PA} - C_T^{Poll}$) are expected to be positive after a 'relatively short' period of physical activity T since $C^{Acc} = 0$ (which would be in line with the current research into health impact assessment).

As a conclusion, we here suggest that the costs of bicycle accidents are likely to be underestimated in current health impact assessment studies. Future research is here encouraged to pay special attention on the impact(s) bicycle accidents may have on the level of physical activity of cyclists (and then on the health benefits of cycling). Moreover, greater attention should be paid on minor/slight injuries as they represent the largest share of bicycle accidents and lead to non-negligible costs for cyclists (see e.g. Rabl and de Nazelle, 2012).

6.5 Concluding words

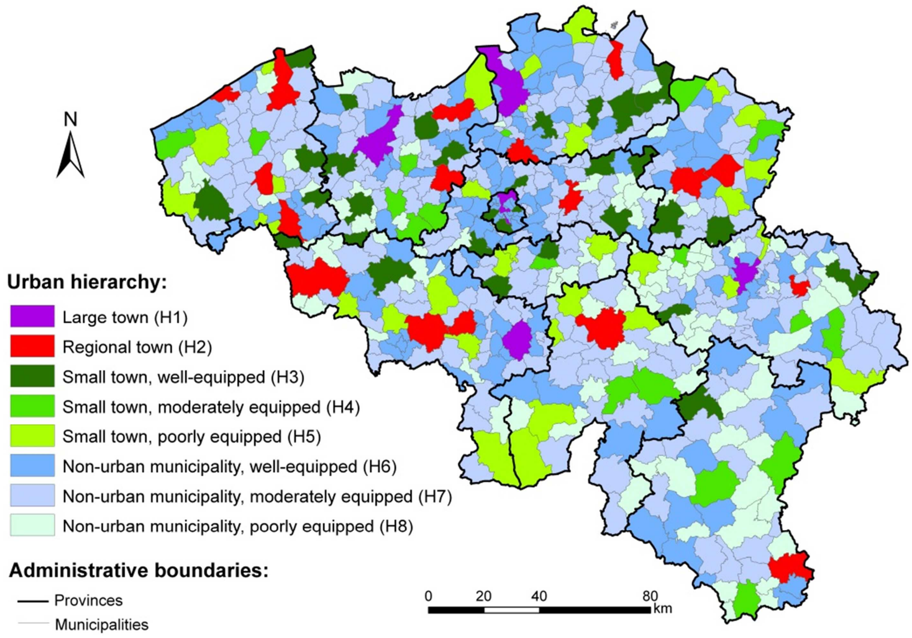
Despite some weaknesses, we believe that this thesis was able to identify some of the main spatial determinants that influence commuter cycling and the risks of being involved in a road accident when cycling. Taking advantage of the recent advances made into the scientific research, we here aimed at delivering findings corrected for a number of statistical biases and resulting in sound recommendations for planners and policy makers. We are also convinced that the present thesis provides a new tool helping planners to prevent road accidents in general. It is now hoped that some of our findings will be taken into consideration by planners and policy makers to support policies aiming at encouraging bicycle use and making it safer…

Appendix A

Notes to Chapter 2

Appendix A.1: The urban hierarchy of Belgian municipalities

(Source: Van Hecke, 1998)



Appendix B

Notes to Chapter 3

Appendix B.1: Variables used: description, units of measurement and data sources

Group	Variable	Description	Units	Source
Dependent variable				
	% cycle commuting (y)	Proportion of commuting by bicycle	Percent	2001 Census
Independent variables				
Demographic data	% working men	Percentage of working people who are men	Percent	DGSEI (2001b)
	% age 1 (< 25)	Percentage of working people who are less than 25 years old	Percent	2001 Census
	% age 2 (45-54)	Percentage of working people who are between 45 and 54 years old	Percent	2001 Census
	% age 3 (> 54)	Percentage of working people who are more than 54 years old	Percent	2001 Census
	% young children (≤ 5 years)	Percentage of working households (i.e. those with one or more working parents) having one or more young children (being less than 5 years old)	Percent	Own computation from 2001 Census

continued on next page

continued

Group	Variable	Description	Units	Source
Socio-economic data	% education 1 (primary school)	Percentage of working people having a primary school as their highest qualification	Percent	2001 Census
	% education 2 (secondary school)	Percentage of working people having a secondary school leaving certificate as their highest qualification	Percent	2001 Census
	% education 3 (university degree)	Percentage of working people having a university (or equivalent) degree as their highest qualification	Percent	2001 Census
	Income	Median income (per capita)	Euro (.10^3)	DGSEI (2001b)
	% bad health	Percentage of inhabitants feeling they have a bad state of health	Percent	2001 Census
	% car owner	Percentage of households that do not own any car	Percent	DGSEI (2001b)
Environmental and policy-related data	Population density	Population density	Inhabitants/km²	DGSEI (2001b)
	Jobs density	Jobs density	Jobs/km²	DGSEI (2001b)
	Commuting distance	Average daily commuting distance of working people, by day	Kilometres	2001 Census
	Town distance	Minimum road distance to the closest town (see Chapter 2 for more details of how this variable is defined)	Kilometres	Vandenbulcke et al. (2007)
	% short cycle commuting	Percentage of commuters who live no more than 10 km from their workplace	Percent	2001 Census
	Town size (urban rank)	Urban hierarchy of Belgian municipalities (from large towns (1) to small villages (8))	1-8	Van Hecke (1998)
	% urban areas	Percentage of the municipality which is urbanised	Percent	DGSEI (2004)
	% forested areas	Percentage of the municipality which is forested	Percent	DGSEI (2004)
	% agricultural areas	Percentage of the municipality which is agricultural	Percent	DGSEI (2004)

continued on next page

continued

Group	Variable	Description	Units	Source
Environmental and policy-related data	% public services areas	Percentage of the surface area of the municipality which is used for public services (e.g. municipal offices, schools)	Percent	DGSEI (2004)
	% recreational areas	Percentage of the surface area of the municipality which is used for recreation (e.g. parks, sport terrains)	Percent	DGSEI (2004)
	Slope	Mean gradient of the municipal road network (excluding motorways and main roads)	Degree	Own computation from EROS data (2002)
	% dissatisfaction of cycling facilities	Percentage of households estimating that they have low-quality cycling facilities located in their neighbourhood	Percent	2001 Census
	Bicycle theft	Average annual number of bicycle thefts	Bicycle thefts	Federal Police (2000-2002)
	Theft risk	Average annual number of bicycle thefts, divided by the total number of cyclists in the municipality	Number of bicycle thefts per cyclist	Own computation from Federal Police data (2000-2002) and 2001 Census
	Accident risk	Average number of cyclists who are victims of accidents per 100,000 minutes spent on a bicycle	Victims (cyclists) per 100,000 minutes	Own computation from DGSEI data (2002-2005) and 2001 Census
	Air quality	Mean concentration of particulate matter (PM10)	Microgram/m³	Own computation from IRCEL-CELINE data (2000-2005)
	Traffic volume 1 (regional roads)	Annual number of vehicle-kilometres (.10^6) per kilometre of regional road	10^6 vehicle-km per kilometre of network	FPS Mobility and Transports (DGSEI, 2000)
	Traffic volume 2 (municipal/local roads)	Annual number of vehicle-kilometres (.10^6) per kilometre of municipal/local road	10^6 vehicle-km per kilometre of network	FPS Mobility and Transports (DGSEI, 2000)

Appendix B.2: Regression coefficients for the spatial regime specification

(ML estimation, with heteroskedasticity correction) – Dependent variable: proportion of cyclists among commuters who travel less than 10 km (in municipality *i*)

	ML, spatial regimes & heterosk. correction	
	North	*South*
Intercept	3.8716**	3.3259***
	[0,0000]	[0,0000]
Lag coefficient (ρ)	0.4740***	0.4740***
	[0,4475]	[0,4475]
Demographic variables		
% working men	0.0239	0.0211
	[0,7353]	[0,6582]
% age 2 (45-54)	-0.0562**	-0.0307***
	[-0,7019]	[-0,4011]
% age 3 (> 54)†	-0.1074	-0.0680
	[-0,1317]	[-0,0867]
% young children	-0.0304**	-0.0292***
(≤ 5 years)	[-0,3243]	[-0,3486]
Socio-economic variables		
% education 3	-0.1098	-0.1839
(university degree)†	[-0,2123]	[-0,3586]
Income	-	-
	-	-
% bad health	-0.0241**	-0.0194**
	[-0,2792]	[-0,2939]
Environmental and policy-related variables		
Commuting distance	0.0103	0.0076*
	[0,1148]	[0,1100]
Town size	-0.0908***	-0.0161
(urban rank)	[-0,3201]	[-0,0587]
Slope†	-0.3655***	-0.3383***
	[-0,1929]	[-0,3002]
% dissatisfaction with	-0.0045**	-0.0072***
cycling facilities	[-0,1277]	[-0,3185]

continued on next page

continued

	ML, spatial regimes & heterosk. correction	
	North	***South***
Accident risk†	-0.6580***	-0.2901***
	[-0,0803]	[-0,0855]
Air pollution	0.0069*	-0.0054
	[0,1136]	[-0,0852]
Traffic volume 2	-0.3684	-0.6670***
(municipal/local)†	[-0,0426]	[-0,0919]
N	589 (N_{North} = 308; N_{South} = 281)	
Log Likelihood	-56.65	
Akaike information criterion (*AIC*)	173.30	
Schwarz information criterion (*SIC*)	304.66	

*Significant at the 90% level; **Significant at the 95% level; ***Significant at the 99% level
Standardised regression coefficients are given in brackets
†: logarithmically transformed variables

Appendix B.3: Impact of spatial interactions

In Section 3.5.5.3., effects of changes in the values of the explanatory variables of the spatial regime regression could be incorrectly interpreted due to the presence of complex spatial interactions in the model. '**Direct effects**' on cycling levels in municipality i may arise from a change in a single explanatory variable in this municipality i; these include: (1) the effect of a change **through** i, and (2) feedback influences resulting from impacts (caused by changes in i) passing through the neighbouring municipalities j, and coming back to i (feedback loop). '**Indirect effects**' (or spillover effects) on cycling levels in i may also arise from changes in explanatory variables in j (LeSage and Fisher, 2008; Fisher et al., 2009; Kirby and LeSage, 2009; LeSage and Pace, 2009).

Given that such direct and indirect effects may affect the validity of the results reported in Table 3.5, we hence checked the magnitude of their impact (on the results) by comparing the parameter estimates of the spatial lag regime specification with scalar summary impact measures (provided by LeSage and

Fisher (2008) and LeSage and Pace (2009))[1]. Discrepancies between parameter estimates and direct impact estimates were observed (Appendix B.3 (see Table below)), indicating that changes in the explanatory variables in municipality *i* produce feedback effects on cycling levels in *i*; such changes in the explanatory variables indeed influence the neighbouring municipalities' cycling levels, which afterwards influence these in municipality *i* (feedback loop). Given that direct impact estimates exceed the parameter estimates in Table 3.5, the feedback effects are positive and hence (very slightly) increase the importance of changes in explanatory variables on cycling levels. However, the difference (between the direct impact estimates and the parameter estimates) is small and suggests that such feedback effects are weak. As a result, the parameter estimates reported in Table 3.5 give a reasonable measure of the direct impact of changes (in explanatory variables) on cycling levels in *i*.

As suggested by Appendix B.3 (Table), changes in the neighbouring municipalities *j* of an explanatory variable also cause indirect effects (or spillover influences) on cycling levels in *i*. It hence confirms previous results exhibited in Table 3.5, suggesting that spillover effects (captured through the spatial autoregressive coefficient) exist between a municipality and its neighbourhood. Such effects are (slightly) larger compared with the direct effects, suggesting that spillover effects should not be ignored in the model. We however do not attempt to interpret the separate contribution of a change in the explanatory variables in *j* to the overall spillover effect. The use of a spatial Durbin model (SDM) is probably more convenient (in terms of inference) for the computation of the impact estimates. Although this specification is likely to suffer from multicollinearity between Wy and WX (Angeriz et al., 2008), it incorporates an additional matrix $W\boldsymbol{\theta}$ ($\boldsymbol{\theta}$: vector of parameters of the SDM associated with the spatially lagged explanatory variables WX) and exhibits a great deal of heterogeneity arising from the presence of this latter matrix $W\boldsymbol{\theta}$ in the total impact estimates (as opposed to the SAR case).

[1] Several pre-released functionalities of the 'spdep' R Package were here used within the framework of this chapter ('impacts.sarlm'; implemented by Roger Bivand), but the results were not reported. Note that sparse matrices were used to estimate the traces of the power series of the spatial weights matrix (10,000 simulated draws); they indeed seem to perform better than powering Monte Carlo simulations (Roger Bivand, personal communication).

	NORTH			SOUTH		
	Direct	*Indirect*	*Total*	*Direct*	*Indirect*	*Total*
Intercept	2.4803**	2.5421**	5.0224**	4.6142***	4.7060***	9.3202***
Demographic variables						
% working men	0.0316***	0.0320***	0.0636***	0.0009	0.0010	0.00194
% age 2 (45-54)	-0.0448***	-0.0457***	-0.0904***	-0.0218***	-0.0222***	-0.0440***
% age 3 (> 54)†	-0.1150	-0.1174	-0.2325	-0.0742	-0.0767	-0.15092
% young children (≤ 5 years)	-0.0391***	-0.0399***	-0.0789***	-0.0265***	-0.0271***	-0.0536***
Socio-economic variables						
% education 3 (university degree)†	-0.1045	-0.1079	-0.2123	-0.3350***	-0.3415***	-0.6765***
Income	0.0333***	0.0341**	0.0674***	-0.0030	-0.0030	-0.00602
% bad health	-0.0105*	-0.0106	-0.0210	-0.0157***	-0.0160**	-0.0317***
Environmental and policy-related variables						
Commuting distance	-0.0177***	-0.0181***	-0.0358***	-0.0050	-0.0051	-0.01013
Town size (urban rank)	-0.1225***	-0.1256***	-0.2482***	-0.0387***	-0.0397***	-0.0784***
Slope†	-0.2063***	-0.2078***	-0.4141***	-0.2110***	-0.2150***	-0.4261***
% dissatisfaction with cycle facilities	-0.0055***	-0.0056***	-0.0112***	-0.0048***	-0.0049***	-0.0098***
Accident risk†	-0.8162***	-0.8350***	-1.6512***	-0.1590***	-0.1632***	-0.3222***
Air pollution	0.0148***	0.0151***	0.0298***	-0.0058	-0.0059	-0.01170
Traffic volume 2 (municipal roads)†	-0.2500	-0.2556	-0.5056	-0.4857***	-0.4974***	-0.9831***

*Significant at the 90% level; **Significant at the 95% level; ***Significant at the 99% level
Total impact estimates are the sum of the direct and indirect impact estimates
†: logarithmically transformed variables

Appendix B.3 (Table): Direct, indirect and total impact estimates (means), based on the spatial lag regime specification

Appendix C

Notes to Chapter 4

Appendix C.1: Infrastructure factors – Description and data sources

	Description	Data source
Bridge	Bridges and elevated roads with safeguards on both side	Own digitization, from BRIC (Brussels UrbIS 2007-2008, GeoLoc) & Google Earth (2004, 2007, 2009)
Tunnel	Tunnels or parts of the road network situated below an elevated infrastructure	Own digitization, from BRIC (Brussels UrbIS 2007-2008, GeoLoc) & Google Earth (2004, 2007, 2009)
Traffic-calming area	Traffic-calming areas. $\Psi = 1$ (30 km/h area), 2 (pedestrian area), 3 (residential area), 4 (all types of traffic-calming areas, i.e. 1-3)	Own digitization, from BRIC (Brussels UrbIS 2007-2008, cycling map (BCR 2006 & 2008), Ministry of the Brussels-Capital Region (IRIS 2), Town of Brussels (Map of the 'comfort area')
Crossroad	Crossroads/intersections. $\Psi = 0$ (no crossroad), 1 (yield/stop signal), 2 (right-of-way), 3 (traffic light), 4 (roundabout), 5 (crossroad with right-turn), 6 (pedestrian light)	Own digitization, from BRIC (Brussels UrbIS 2007-2008, GeoLoc), Google Earth (2004, 2007, 2009)
Tram tracks	Tram tracks. $\Psi = 0$ (no tram track), 1 (crossing tram tracks), 2 (tram tracks in crossable reserved lanes), 3 (on-road tram tracks)	Own digitization, from BRIC (Brussels UrbIS 2007-2008, GeoLoc), Google Earth (2004, 2007, 2009), STIB-MIVB / BRSI

continued on next page

continued

	Description	Data source
Cycle facility	Cycle facilities. $\Psi = 0$ (no cycle facility), 1 (unidirectional separated cycle lane), 2 (bidirectional separated cycle lane), 3 (marked cycle lane), 4 (suggested cycle lane) or 5 (bus and bicycle lane)	Own digitization, from DGSEI (2006-2008), BRIC (Brussels UrbIS 2007-2008, GeoLoc, cycling map (BCR 2006 & 2008), Google Earth (2004, 2007, 2009)
Parking area (aspect-based)	Parking areas (aspect-based). $\Psi = 0$ (no parking area), 1 (longitudinal parking), 2 (head-in angle parking), 3 (back-in angle parking), 4 (parking perpendicular to the road) or 5 (other type of parking area)	Own digitization, from DGSEI (2006-2008), BRIC (Brussels UrbIS 2007-2008, GeoLoc), Google Earth (2004, 2007, 2009)
Contraflow cycling	Streets where contraflow cycling is permitted	Own digitization, from BRIC (Brussels UrbIS 2007-2008, GeoLoc, cycling map (BCR 2006 & 2008), OneWayMap application), Google Earth (2004, 2007, 2009)
Discontinuity	Discontinuities in the cycle facilities (i.e. locations where a cycle facility is disrupted)	Own digitization, from BRIC (Brussels UrbIS 2007-2008, GeoLoc, cycling map (BCR 2006 & 2008), Google Earth (2004, 2007, 2009)
Parking area (function-based)	Parking areas (function-based). $\Psi = 1$ (park-and-ride, public or private parking area), 2 (delivery parking), 3 (diplomatic corps parking), 4 (disabled parking), 5 (taxi parking), 6 (all types of parkings, i.e. 1-5)	BRIC (Brussels UrbIS 2007-2008)
Public transport	Public transport stops. $\Psi = 1$ (bus stop), 2 (tram stop), 3 (all types of public transport stops, i.e. 1-2)	BRIC (Brussels UrbIS 2007-2008)
Public administration	Administrative buildings. $\Psi = 1$ (european administrative building), 2 (regional administrative building), 3 (all types of administrative buildings, i.e. 1-2)	BRIC (Brussels UrbIS 2007-2008)

Continued on next page

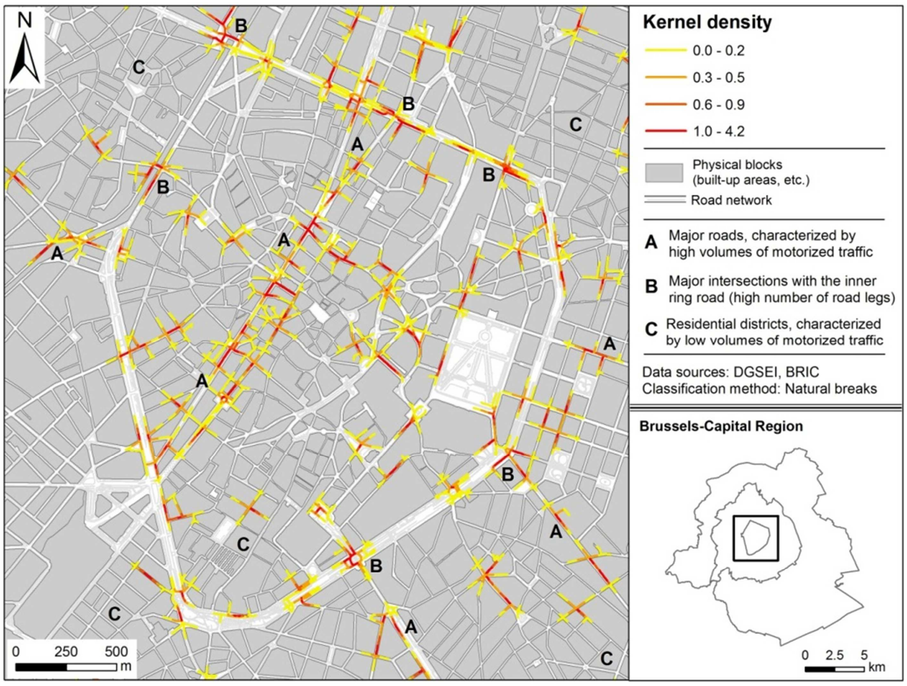
continued

	Description	Data source
School	Schools. $\Psi = 1$ (primary or secondary school), 2 (international primary or secondary school), 3 (superior school), 4 (all types of schools, i.e. 1-3)	BRIC (Brussels UrbIS 2007-2008)
Industrial estate	Industrial estates, sites of economic activities	BRIC (Brussels UrbIS 2007-2008)
Shopping center	Shopping centers / malls, and shopping arcades	BRIC (Brussels UrbIS 2007-2008)
Supermarket	Food and home improvement superstores	BRIC (Brussels UrbIS 2007-2008)
Service station	Service stations / petrol pumps	BRIC (Brussels UrbIS 2007-2008)
Cultural building	Cultural centres, museums, theatres, cinemas, auditoriums, etc.	BRIC (Brussels UrbIS 2007-2008)
Sports complex	Sports complexes	BRIC (Brussels UrbIS 2007-2008)
Playground	Playgrounds	BRIC (Brussels UrbIS 2007-2008)
Religious building	Religious buildings. $\Psi = 1$ (synagogue), 2 (protestant church), 3 (orthodox church), 4 (mosque), 5 (catholic buildings), 6 (all types of religious buildings, i.e. 1-5)	BRIC (Brussels UrbIS 2007-2008)
Police building	Police stations and departments	BRIC (Brussels UrbIS 2007-2008)
Hospital	Hospitals, clinics and health centres	BRIC (Brussels UrbIS 2007-2008)
Embassy	Embassies	BRIC (Brussels UrbIS 2007-2008)

Ψ: Nominal variable, taking on different values for each infrastructure variable (one value = one kind of infrastructure or facility)

Appendix C.2: Blackspots of cycling accidents in the Pentagon (2006-2008)

Network kernel densities (equal split discontinuous function). Bandwidth: 100m; cell width: 10m.



Although several bandwidths were experienced to examine the variation in the density values along the network, a 100m value is here selected since it seems to provide a more adequate representation of the black spots for cyclists (as well as for other 'slow road users', such as pedestrians). Such a choice is also justified by the fact that 100-300m bandwidths are commonly used in urban studies modelling pedestrian catchment areas and accidents, at the scale of neighbourhoods (300m), blocks (200m) and streets (100m) (Cervero, 1998; Frey, 1999; Calthorpe and Fulton, 2001; Cervero, 2004; Okabe et al., 2009; Porta et al., 2009; Dai et al., 2010). In order to avoid edge effects, cycling accidents located outside the study region (i.e. in the Flemish municipalities) are also considered when applying the equal split method in SANET. The densities obtained using such a method are finally manually corrected with the aim to account for the presence of road elevations along the network (e.g. bridges, tunnels, etc.), given that SANET ignores these latter when computing the densities.

Appendix D

Notes to Chapter 5

Appendix D.1: List of risk factors

Description, categorical values (Ψ), units and data sources

Variable	Definition	Ψ values	Units	Data source
Infrastructure				
Bridge[a]	1 if the accident/control occurred on a bridge (with safeguards on both sides), 0 otherwise	-	-	Own digitalization and computation, from BRIC (Brussels UrbIS 2007-2008, GeoLoc) & Google Earth (2004, 2007, 2009)
Tunnel[a]	1 if the accident/control occurred in a tunnel or below an elevated infrastructure, 0 otherwise	-	-	Own digitalization and computation, from BRIC (Brussels UrbIS 2007-2008, GeoLoc) & Google Earth (2004, 2007, 2009)
Traffic-calming area Ψ[a]	1 if the accident/control occurred in a type Ψ traffic-calming area, 0 otherwise	Ψ = 1 (30 km/h area), 2 (pedestrian area), 3 (residential area), 4 (all types of traffic-calming areas, i.e. 1-3)	-	Own digitalization and computation, from BRIC (Brussels UrbIS 2007-2008, cycling map (BCR 2006 & 2008), Ministry of the Brussels-Capital Region (IRIS 2), City of Brussels (Map of the "comfort area")
Crossroad Ψ[a]	1 if the accident/control occurred in a type Ψ crossroad, 0 otherwise	Ψ = 0 (no crossroad), 1 (yield/stop signal), 2 (right-of-way), 3 (traffic light), 4 (roundabout), 5 (crossroad with right-turn), 6 (pedestrian light)	-	Own digitalization and computation, from BRIC (Brussels UrbIS 2007-2008, GeoLoc), Google Earth (2004, 2007, 2009)

continued on next page

continued

Variable	Definition	Ψ values	Units	Data source
Complexity index Ψ	Complexity index at the place of the accident/control, with Ψ bandwidth (m)	Ψ = 10, 20, 30, 40, 50, 75 or 100 m	Meters	Own computation, from BRIC (Brussels UrbIS)
Tram tracks Ψ[a,b]	1 if the accident/control occurred on or close to a type Ψ tram track infrastructure, 0 otherwise	Ψ = 0 (no tram track), 1 (crossing tram tracks), 2 (tram tracks in crossable reserved lanes), 3 (on-road tram tracks)	-	Own digitalization and computation, from BRIC (Brussels UrbIS 2007-2008, GeoLoc), Google Earth (2004, 2007, 2009), STIB-MIVB / BRSI
Cycle facility Ψ[a,b]	1 if the accident/control occurred on a type Ψ cycle facility, 0 otherwise	Ψ = 0 (no cycle facility), 1 (unidirectional separated cycle lane), 2 (bidirectional separated cycle lane), 3 (marked cycle lane), 4 (suggested cycle lane) or 5 (bus and bicycle lane)	-	Own digitalization and computation, from FPS Economy (2006-2008), BRIC (Brussels UrbIS 2007-2008, GeoLoc, cycling map (BCR 2006 & 2008), Google Earth (2004, 2007, 2009)
Parking area Ψ[a,b]	1 if the accident/control occurred close to a type Ψ parking area, 0 otherwise	Ψ = 0 (no parking area), 1 (longitudinal parking), 2 (head-in angle parking), 3 (back-in angle parking), 4 (parking perpendicular to the road) or 5 (other type of parking area)	-	Own digitalization and computation, from FPS Economy (2006-2008), BRIC (Brussels UrbIS 2007-2008, GeoLoc), Google Earth (2004, 2007, 2009)
Proximity parking-cycle facility Ψ[a,b]	1 if the accident/control occurred on a type Ψ cycle facility, very close to a parking area (d ≤ 0.8 m, and outside a crossroad), 0 otherwise	Ψ = 1 (unidirectional separated cycle lane), 2 (bidirectional separated cycle lane), 3 (marked cycle lane), 4 (suggested cycle lane) or 5 (bus and bicycle lane), 6 (all types of cycle facilities, i.e. 1-5)	-	Own digitalization and computation, from NIS-FPS Economy (2006-2008), BRIC (Brussels UrbIS 2007-2008, GeoLoc, cycling map (BCR 2006 & 2008), Google Earth (2004, 2007, 2009)
Contraflow cycling[a,b]	1 if the accident/control occurred in a contraflow cycling and in the opposite direction of motorised vehicles (i.e. in the direction of the contraflow), 0 otherwise	-	-	Own digitalization and computation, from BRIC (Brussels UrbIS 2007-2008, GeoLoc, cycling map (BCR 2006 & 2008), OneWayMap application), Google Earth (2004, 2007, 2009)
Major road	1 if the accident/control occurred on a major road, 0 otherwise	-	-	Own computation, from BRIC (Brussels UrbIS 2007-2008)

continued on next page

continued

Variable	Definition	Ψ values	Units	Data source
Number of garages Ψ (≤ 100m)	Number of garages (in a range Ψ) over a network distance ≤ 100m from the place of the accident/control	Ψ = 0, 0-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, > 70 garage(s)	-	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Garage length	Sum of all the garage lengths over a network distance ≤ 100m from the place of the accident/control	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Garage ≤ Ψ (m)	1 if the accident/control occurred over a network distance d ≤ Ψ (m) from a garage, 0 otherwise	Ψ = 10, 50 or 100 m	-	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance garage	Network distance to the closest garage	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance crossroad	Network distance to the closest crossroad, whatever the type of crossroad	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance discontinuity[a,b]	Network distance to the closest discontinuity (on cycle facilities)	-	Meters	Own digitalization and computation, from BRIC (Brussels UrbIS 2007-2008, GeoLoc, cycling map (BCR 2006 & 2008), Google Earth (2004, 2007, 2009)
Distance city centre	Network distance to the Brussels' town hall (city centre)	-	Meters	Own digitalization and computation, from Google Map/Earth 2009
Distance major road	Network distance to the closest crossroad of a major road	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance parking area Ψ	Network distance to the closest type Ψ parking area	Ψ = 1 (park-and-ride, public or private parking area), 2 (delivery parking), 3 (diplomatic corps parking), 4 (disabled parking), 5 (taxi parking), 6 (all types of parkings, i.e. 1-5)	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance public transport Ψ	Network distance to the closest type Ψ public transport stop	Ψ = 1 (bus stop), 2 (tram stop), 3 (all types of public transport stops, i.e. 1-2)	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)

continued on next page

continued

Variable	Definition	Ψ values	Units	Data source
Distance public administration Ψ	Network distance to the closest type Ψ administrative building	$\Psi = 1$ (european administrative building), 2 (regional administrative building), 3 (all types of administrative buildings, i.e. 1-2)	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance school Ψ	Network distance to the closest type Ψ school	$\Psi = 1$ (primary or secondary school), 2 (international primary or secondary school), 3 (superior school), 4 (all types of schools, i.e. 1-3)	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance industrial estate	Network distance to the closest industrial estate	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance shopping center	Network distance to the closest shopping center / mall	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance supermarket	Network distance to the closest supermarket	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance service station	Network distance to the closest service station / petrol pump	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance cultural building	Network distance to the closest cultural building / center	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance sports complex	Network distance to the closest sports complex	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance playground	Network distance to the closest playground	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance religious building Ψ	Network distance to the closest type Ψ religious building	$\Psi = 1$ (synagogue), 2 (protestant church), 3 (orthodox church), 4 (mosque), 5 (catholic buildings), 6 (all types of religious buildings, i.e. 1-5)	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance police building	Network distance to the closest police building	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)

continued on next page

continued

Variable	Definition	Ψ values	Units	Data source
Distance hospital	Network distance to the closest hospital	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Distance embassy	Network distance to the closest embassy	-	Meters	Own computation, from BRIC (Brussels UrbIS 2007-2008)
Traffic				
Car traffic $\Psi^{a,b}$ (06:00 a.m. - 10:59 p.m.)	1 if the accident/control occurred on a road with intensity Ψ car traffic between 06:00 a.m. and 10:59 p.m., 0 otherwise	Ψ = 1, 2, 3, 4, 5 (class 1 = very low car traffic ; class 5 = very high car traffic)	-	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007-2008)
Car traffic $\Psi^{a,b}$ (08:00 a.m. - 08:59 a.m.)	1 if the accident/control occurred on a road with intensity Ψ car traffic between 08:00 a.m. and 08:59 a.m., 0 otherwise	Ψ = 1, 2, 3, 4, 5 (class 1 = very low car traffic ; class 5 = very high car traffic)	-	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007-2008)
Car traffic $\Psi^{a,b}$ (5:00 p.m. - 5:59 p.m.)	1 if the accident/control occurred on a road with intensity Ψ car traffic between 5:00 p.m. and 5:59 p.m., 0 otherwise	Ψ = 1, 2, 3, 4, 5 (class 1 = very low car traffic ; class 5 = very high car traffic)	-	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007-2008)
Van traffic $\Psi^{a,b}$ (06:00 a.m. - 10:59 p.m.)	1 if the accident/control occurred on a road with intensity Ψ van traffic between 06:00 a.m. and 10:59 p.m., 0 otherwise	Ψ = 1, 2, 3, 4, 5 (class 1 = very low van traffic ; class 5 = very high van traffic)	-	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007-2008)
Van traffic $\Psi^{a,b}$ (08:00 a.m. - 08:59 a.m.)	1 if the accident/control occurred on a road with intensity Ψ van traffic between 08:00 a.m. and 08:59 a.m., 0 otherwise	Ψ = 1, 2, 3, 4, 5 (class 1 = very low van traffic ; class 5 = very high van traffic)	-	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007-2008)
Van traffic $\Psi^{a,b}$ (5:00 p.m.- 5:59 p.m.)	1 if the accident/control occurred on a road with intensity Ψ van traffic between 5:00 p.m. and 5:59 p.m., 0 otherwise	Ψ = 1, 2, 3, 4, 5 (class 1 = very low van traffic ; class 5 = very high van traffic)	-	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007-2008)

continued on next page

continued

Variable	Definition	Ψ values	Units	Data source
Lorry/truck traffic Ψ[a,b] (06:00 a.m.-10:59 p.m.)	1 if the accident/control occurred on a road with intensity Ψ truck traffic between 06:00 a.m. and 10:59 p.m., 0 otherwise	Ψ = 1, 2, 3, 4, 5 (class 1 = very low truck traffic ; class 5 = very high truck traffic)	-	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007-2008)
Lorry/truck traffic Ψ[a,b] (08:00 a.m.-08:59 a.m.)	1 if the accident/control occurred on a road with intensity Ψ truck traffic between 08:00 a.m. and 08:59 a.m., 0 otherwise	Ψ = 1, 2, 3, 4, 5 (class 1 = very low truck traffic ; class 5 = very high truck traffic)	-	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007-2008)
Lorry/truck traffic Ψ[a,b] (5:00 p.m.- 5:59 p.m.)	1 if the accident/control occurred on a road with intensity Ψ truck traffic between 5:00 p.m. and 5:59 p.m., 0 otherwise	Ψ = 1, 2, 3, 4, 5 (class 1 = very low truck traffic ; class 5 = very high truck traffic)	-	Own computation, from STRATEC/IBGE-BIM (2006), BRIC (Brussels UrbIS 2007-2008)
Environment				
Slope	Maximum slope (to neighbouring pixels) computed at the pixel where the accident/control took place	-	Degree	Own computation, from EROS (2002)
Green areas ≤ Ψ (m)	1 if the accident/control occurred over an euclidean distance d ≤ Ψ (m) from a green area, 0 otherwise	Ψ = 10, 20, 30, 40 or 50 m	-	Own computation, from BRIC (Brussels UrbIS, 2007-2008)

[a] Year is controlled

[b] Direction of travel is controlled

Appendix D.2: Descriptive statistics of the selected (discrete) risk factors

Variable	*Ψ*	% No Acc	% Acc	χ^2 test (p)	*F* test (p)	OR	(LCI-UCI)	$P_{Acc > Abs}$	Risk
Infrastructure									
Bridge	-	0.6	**2.0**	0.00	0.00	3.53	(1.53-6.93)	1.00	▶▶▶
Tunnel†	-	**0.2**	0.0	**0.48**	**0.61**	0.68	(0.01-2.80)	0.21	◀▶
Traffic-calming area	1	**11.4**	9.2	**0.14**	**0.13**	0.80	(0.58-1.06)	0.06	◀
	2	**0.8**	0.5	**0.64**	**0.60**	0.85	(0.21-2.02)	0.30	◀▶
	3	**0.4**	0.3	**1.00**	**1.00**	1.21	(0.21-3.36)	0.50	◀▶
	4	**12.6**	10.0	**0.10**	**0.09**	0.78	(0.58-1.03)	0.04	◀◀
Crossroad	0	**82.2**	44.3	0.00	0.00	0.17	(0.14-0.21)	0.00	◀◀◀
	1	2.2	**11.0**	0.00	0.00	5.71	(3.86-8.19)	1.00	▶▶▶
	2	10.2	**18.5**	0.00	0.00	2.01	(1.56-2.55)	1.00	▶▶▶
	3	3.3	**17.7**	0.00	0.00	6.41	(4.67-8.60)	1.00	▶▶▶
	4	1.3	**6.7**	0.00	0.00	5.63	(3.41-8.83)	1.00	▶▶▶
	5	0.8	**1.5**	**0.20**	**0.16**	2.03	(0.85-3.99)	0.94	▶
	6†	0.1	**0.3**	**0.38**	**0.18**	6.10	(0.69-23.73)	0.94	▶
Tram tracks	0	**95.4**	82.5	0.00	0.00	0.23	(0.17-0.30)	0.00	◀◀◀
	1	0.5	**5.7**	0.00	0.00	13.52	(6.66-25.43)	1.00	▶▶▶
	2	1.3	**3.7**	0.00	0.00	3.07	(1.69-5.11)	1.00	▶▶▶
	3	2.8	**8.2**	0.00	0.00	3.18	(2.13-4.55)	1.00	▶▶▶
Cycle facility	0	**93.0**	81.0	0.00	0.00	0.32	(0.25-0.42)	0.00	◀◀◀
	1	1.4	**5.0**	0.00	0.00	3.70	(2.18-5.86)	1.00	▶▶▶
	2	2.2	**3.7**	**0.06**	**0.06**	1.74	(1.01-2.76)	0.98	▶▶

continued on next page

continued

Variable	Ψ	% No Acc	% Acc	χ^2 test (p)	F test (p)	OR	(LCI-UCI)	$P_{Acc > Abs}$	Risk
Cycle facility	3	2.2	**7.0**	0.00	0.00	3.44	(2.22-5.08)	1.00	▶▶▶
	4	0.9	**2.5**	0.00	0.00	2.98	(1.45-5.40)	1.00	▶▶▶
	5†	0.2	**0.8**	0.05	0.03	4.86	(1.25-13.21)	0.99	▶▶
Parking area	0	34.1	**58.0**	0.00	0.00	2.67	(2.22-3.20)	1.00	▶▶▶
	1	**63.0**	40.8	0.00	0.00	0.41	(0.34-0.49)	0.00	◀◀◀
	2	**0.7**	0.3	**0.52**	**0.55**	0.75	(0.14-1.99)	0.24	◀▶
	3†	0.1	**0.2**	**1.00**	**0.49**	4.06	(0.29-17.00)	0.82	◀▶
	4	**1.5**	0.3	0.03	0.01	0.32	(0.06-0.81)	0.01	◀◀◀
	5†	**0.5**	0.3	**0.85**	**1.00**	1.00	(0.18-2.73)	0.40	◀▶
Proximity parking-cycle facility	1	0.6	**1.7**	0.02	0.02	3.20	(1.32-6.47)	1.00	▶▶▶
	2†	0.1	**0.3**	**0.38**	**0.18**	6.10	(0.69-23.77)	0.94	▶
	3	1.4	**1.7**	**0.72**	**0.57**	1.35	(0.61-2.47)	0.75	◀▶
	4	0.6	**0.8**	**0.77**	**0.57**	1.61	(0.51-3.60)	0.77	◀▶
	5†	0.0	0.0	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	6	2.6	**4.5**	0.03	0.02	1.80	(1.10-2.74)	0.99	▶▶▶
Contraflow cycling	-	**10.3**	5.3	0.00	0.00	0.51	(0.34-0.72)	0.00	◀◀◀
Major road	-	12.9	46.7	0.00	0.00	5.95	(4.86-7.22)	1.00	▶▶▶
Number of garages (≤ 100m)	0	**6.3**	4.5	**0.13**	**0.12**	0.73	(0.47-1.08)	0.05	◀
	0-10	41.9	**47.7**	0.01	0.01	1.27	(1.06-1.51)	0.99	▶▶▶
	11-20.	32.0	**35.0**	**0.17**	**0.16**	1.15	(0.95-1.39)	0.92	▶
	21-30	**15.8**	8.8	0.00	0.00	0.53	(0.38-0.70)	0.00	◀◀◀

continued on next page

continued

Variable	Ψ	% No Acc	% Acc	χ^2 test (p)	F test (p)	OR	(LCI-UCI)	$P_{Acc > Abs}$	Risk
Number of garages (≤ 100m)	31-40	**6.2**	4.0	0.02	0.02	0.66	(0.41-0.98)	0.02	◀◀
	41-50	**2.8**	2.7	**0.85**	**0.88**	1.02	(0.56-1.67)	0.48	◀▶
	51-60	0.9	**1.5**	**0.29**	**0.25**	1.84	(0.78-3.59)	0.91	▶
	61-70†	**0.3**	0.2	**0.94**	**1.00**	1.15	(0.12-3.78)	0.43	◀▶
	> 70†	0.1	**0.2**	**1.00**	**0.59**	2.68	(0.23-10.30)	0.74	◀▶
Garage ≤ Ψ	≤ 10 m	**48.2**	30.2	0.00	0.00	0.47	(0.38-0.56)	0.00	◀◀◀
	≤ 50 m	**84.6**	80.3	0.01	0.01	0.75	(0.59-0.94)	0.01	◀◀◀
	≤ 100 m	93.8	**95.5**	**0.13**	**0.12**	1.43	(0.93-2.14)	0.95	▶
Traffic									
Car traffic (06:00 a.m. to 10:59 p.m.)	1	**76.4**	41.0	0.00	0.00	0.22	(0.18-0.26)	0.00	◀◀◀
	2	15.2	**31.7**	0.00	0.00	2.59	(2.10-3.17)	1.00	▶▶▶
	3	6.3	**18.2**	0.00	0.00	3.32	(2.52-4.28)	1.00	▶▶▶
	4	2.0	**7.2**	0.00	0.00	3.90	(2.50-5.79)	1.00	▶▶▶
	5	0.1	**2.0**	0.00	0.00	26.82	(6.02-90.67)	1.00	▶▶▶
Van traffic (06:00 a.m. to 10:59 p.m.)	1	**76.1**	38.3	0.00	0.00	0.20	(0.16-0.24)	0.00	◀◀◀
	2	14.7	**32.2**	0.00	0.00	2.76	(2.23-3.37)	1.00	▶▶▶
	3	6.8	**19.3**	0.00	0.00	3.30	(2.53-4.23)	1.00	▶▶▶
	4	1.9	**6.7**	0.00	0.00	3.69	(2.35-5.53)	1.00	▶▶▶
	5	0.4	**3.5**	0.00	0.00	9.14	(4.11-18.19)	1.00	▶▶▶
Truck traffic (06:00 a.m. to 10:59 p.m.)	1	**75.7**	40.2	0.00	0.00	0.22	(0.18-0.26)	0.00	◀◀◀
	2	12.5	**22.0**	0.00	0.00	2.00	(1.58-2.49)	1.00	▶▶▶

continued on next page

continued

Variable	Ψ	% No Acc	% Acc	χ^2 test (p)	F test (p)	OR	(LCI-UCI)	$P_{Acc > Abs}$	Risk
Truck traffic (06:00 a.m. to 10:59 p.m.)	3	8.2	**22.0**	0.00	0.00	3.19	(2.48-4.02)	1.00	▶▶▶
	4	3.1	**11.0**	0.00	0.00	3.92	(2.73-5.44)	1.00	▶▶▶
	5	0.5	**4.8**	0.00	0.00	9.71	(4.88-17.84)	1.00	▶▶▶
Environment									
Green areas ≤ Ψ	≤ 10 m	**9.9**	8.7	**0.39**	**0.40**	0.88	(0.63-1.18)	0.18	◀▶
	≤ 20 m	13.4	**14.3**	**0.60**	**0.55**	1.09	(0.84-1.40)	0.73	◀▶
	≤ 30 m	15.6	**18.5**	**0.10**	**0.09**	1.24	(0.97-1.55)	0.96	▶▶
	≤ 40 m	17.3	**20.7**	**0.07**	**0.07**	1.25	(0.99-1.55)	0.97	▶▶
	≤ 50 m	19.4	**22.2**	**0.14**	**0.14**	1.20	(0.95-1.48)	0.94	▶

† Less than 10 observations for both accidents and controls; care must be taken when analysing the corresponding data

n.s.: not significant at the 90% level; % Acc: proportion of accidents (**bold**: % Acc > % No Acc); % No Acc: proportion of controls (**bold**: % No Acc > % Acc)

Frequentist framework:

χ^2 test (p): p-value of the Chi-Square adjusted test for independence (**bold**: independence not rejected)

F test (p): p-value of the Fisher's exact test for independence (**bold**: independence not rejected)

Bayesian framework:

OR: Odds Ratio; LCI: Lower Credible Interval of the OR (2.5%); UCI: Upper credible interval of the OR (97.5%)

$P_{Acc>Abs}$: Probability that the proportion of accidents is higher (compared with the proportion of controls) when a specific risk factor is present, i.e. when $x_i = 1$

Risk = $P_{Acc>Abs}$: risk of having an accident (for a cyclist) when a specific risk factor x_i is present

Burn-in = 6000 iterations; Post-Burn-in = 20,000 iterations; Number of Markov chains = 3; $\hat{R}$ = 1 (Gelman-Rubin diagnostic) for all variables; MC error < 5% Standard Deviation for all variables; No autocorrelation issue detected.

Symbols for risk: ▶▶▶ Very high (≥ 0.99)**;** ▶▶ Quite high (≥ 0.95)**;** ▶ High ≥ 0.90**;** ◀▶ Moderate (> 0.10 and < 0.90)**;** ◀ Low (≤ 0.10)**;** ◀◀ Quite low (≤ 0.05); ◀◀◀ Very low (≤ 0.01)

Appendix D.3: Descriptive statistics for the continuous risk factors

Variable	Ψ	NO ACCIDENT		ACCIDENT		Wilcoxon test (p)[a]	$P_{Acc > Abs}$[b]	Risk
		Mean[a]	Std. Dev.[a]	Mean[a]	Std. Dev.[a]			
Infrastructure								
Complexity index	10 m	21.3	3.6	28.6	9.7	**0.00**	*1.00*	▶▶▶
	20 m	45.3	10.1	60.0	21.0	**0.00**	*1.00*	▶▶▶
	30 m	72.4	19.4	96.0	36.0	**0.00**	*1.00*	▶▶▶
	40 m	102.7	30.8	136.5	53.1	**0.00**	*1.00*	▶▶▶
	50 m	136.4	43.7	180.7	71.7	**0.00**	*1.00*	▶▶▶
	75 m	235.4	81.2	310.3	123.1	**0.00**	*1.00*	▶▶▶
	100 m	356.4	128.4	469.0	183.1	**0.00**	*1.00*	▶▶▶
Garage length	-	55.4	44.3	52.4	44.9	**0.03**	0.07	◀
Distance garage	-	34.3	93.4	36.5	103.7	**0.00**	*0.69*	◀▶
Distance crossroad	-	51.2	57.0	25.0	39.5	**0.00**	*0.00*	▶▶▶
Distance discontinuity	-	419.6	302.9	356.4	337.2	**0.00**	*0.00*	▶▶▶
Distance city centre	-	4216.8	1906.9	3906.1	2069.4	**0.00**	*0.00*	▶▶▶
Distance major road	-	242.8	230.8	143.0	211.5	**0.00**	*0.00*	▶▶▶
Distance parking area	1	702.9	440.3	629.5	461.6	**0.00**	0.00	▶▶▶
	2	486.2	462.3	407.0	473.2	**0.00**	0.00	▶▶▶
	3	1174.3	1026.8	915.0	908.2	**0.00**	*0.00*	▶▶▶
	4	225.8	223.8	205.3	252.3	**0.00**	*0.03*	▶▶

continued on next page

continued

Variable	Ψ	NO ACCIDENT		ACCIDENT		Wilcoxon test (p)[a]	$P_{Acc > Abs}$[b]	Risk
		Mean[a]	Std. Dev.[a]	Mean[a]	Std. Dev.[a]			
	5	731.6	581.1	639.4	590.4	**0.00**	0.00	▶▶▶
	6	171.7	186.5	142.6	191.0	**0.00**	0.00	▶▶▶
Distance public transport	1	453.3	338.3	360.9	327.2	**0.00**	0.00	▶▶▶
	2	760.4	575.6	683.3	602.2	**0.00**	0.00	▶▶▶
	3	378.4	282.9	283.1	266.8	**0.00**	0.00	▶▶▶
Distance public administration	1	2564.1	1776.7	2170.2	1617.6	**0.00**	*0.00*	▶▶▶
	2	2243.5	1421.0	1774.8	1311.2	**0.00**	*0.00*	▶▶▶
	3	1791.5	1308.3	1458.8	1180.5	**0.00**	*0.00*	▶▶▶
Distance school	1	390.1	248.8	389.6	265.6	0.82	*0.48*	◀▶
	2	2230.9	1604.7	1884.2	1456.3	**0.00**	*0.00*	▶▶▶
	3	1124.6	803.8	938.7	820.9	**0.00**	0.00	▶▶▶
	4	359.0	236.3	335.3	250.3	**0.01**	0.01	▶▶
Distance industrial estate	-	1800.5	942.8	1780.6	955.8	0.85	0.30	◀▶
Distance shopping center	-	2074.7	1282.7	1723.1	1297.8	**0.00**	0.00	▶▶▶
Distance supermarket	-	771.1	538.6	754.4	629.7	**0.02**	*0.27*	◀▶
Distance service station	-	554.3	338.3	539.7	334.3	0.47	0.17	◀▶
Distance cultural building	-	696.2	501.6	611.3	516.3	**0.00**	0.00	▶▶▶
Distance sports complex	-	1128.8	559.6	1119.5	547.8	0.74	0.34	◀▶
Distance playground	-	626.1	359.3	618.8	373.3	0.69	0.33	◀▶

continued on next page

continued

Variable	Ψ	NO ACCIDENT		ACCIDENT		Wilcoxon test (p)[a]	$P_{Acc > Abs}$[b]	Risk
		Mean[a]	Std. Dev.[a]	Mean[a]	Std. Dev.[a]			
Distance religious building	1	2905.3	1898.3	2775.5	1825.9	0.17	0.05	►►
	2	896.1	672.5	2775.5	1825.9	**0.00**	*1.00*	◄◄◄
	3	2116.5	1440.9	1767.8	1388.5	**0.00**	0.00	►►►
	4	1598.1	1254.3	1416.3	1200.8	**0.00**	0.00	►►►
	5	555.0	301.9	530.5	319.1	**0.02**	0.04	►►
	6	458.7	301.2	411.5	310.0	**0.00**	0.00	►►►
Distance police building	-	886.1	509.7	850.3	520.1	**0.03**	0.06	►
Distance hospital	-	1385.0	965.0	1197.9	850.7	**0.00**	*0.00*	►►►
Distance embassy	-	1266.4	1037.9	1031.0	973.4	**0.00**	0.00	►►►
Environment								
Slope	-	2.8	1.7	2.6	1.6	0.14	0.05	◄

Italic: inequal variances; Std. Dev.: Standard Deviation

[a] **Frequentist framework**:

Wilcoxon test (p): p-value of Wilcoxon Rank-Sum test (Mann-Whitney). Significant difference in **bold**

[b] **Bayesian framework**:

$P_{Acc>Abs}$: Probability that the posterior mean of variable x_i is higher for accidents (compared with controls/absences of accidents)

Risk ≈ $P_{Acc>Abs}$: risk of having an accident (for a cyclist) when a specific risk factor x_i is high (e.g. for complexity) or close (for distance-based variables)

Burn-in = 6000 iterations; Post-Burn-in = 20,000 iterations; Number of Markov chains = 3; $\hat{R}$ = 1 (Gelman-Rubin diagnostic) for all variables; MC error < 5% Std.Dev. for all variables; No autocorrelation issue detected.

Symbols for risk: ►►► Very high (≥ 0.99)**; ►►** Quite high (≥ 0.95)**; ►** High ≥ 0.90**; ◄►** Moderate (> 0.10 and < 0.90)**; ◄** Low (≤ 0.10)**; ◄◄** Quite low (≤0.05)**; ◄◄◄** Very low (≤ 0.01)

Appendix D.4: Logistic model – Results from the frequentist framework

Variables	Estimate	SD	Wald Z	OR (e^{β})
Intercept	-5.91***	0.25	-23.54	0.00
Infrastructure				
Complexity index				
Bandwidth = 10 m	0.15***	0.01	15.92	1.16
Bridge & no cycle facility	0.88	0.61	1.43	2.40
Contraflow cycling & no crossroad	-0.65*	0.33	-1.97	0.52
Cycle facility & crossroad				
Facility 1 (unidir.) & Crossroad 1 (yield/stop)	2.05*	0.92	2.24	7.75
Facility 2 (bidir.l) & Crossroad 1 (yield/stop)	2.43*	0.99	2.46	11.36
Facility 3 (mark.) & Crossroad 3 (traffic light)	1.75	1.04	1.68	5.73
Facility 3 (mark.) & Crossroad 4 (roundabout)	2.39**	0.82	2.91	10.96
Facility 4 (sugg.) & Crossroad 2 (right-of-w.)	2.68	1.59	1.69	14.55
Facility 0 (no facility) & Crossroad 4 (round.)	1.02***	0.29	3.51	2.79
Facility 3 (mark.) & Crossroad 0 (no crossr.)	0.74*	0.34	2.19	2.10
Tram tracks				
Class 1 (crossing)	0.84*	0.41	2.06	2.32
Class 2 (crossable reserved lanes)	0.84*	0.36	2.34	2.31
Class 3 (on-road tracks)	1.05***	0.25	4.28	2.87
Number of garages (d ≤100m)				
Range 0 (no garage)	-0.58*	0.28	-2.11	0.56
Distance public administration[a]				
Public administration 2 (regional)	1.07***	0.22	4.86	2.92
Proximity parking-cycle facility				
Parking & Facility 1 (unidirectional)	1.28**	0.45	2.86	3.61
Parking & Facility 2 (bidirectional)	2.06*	0.96	2.14	7.86
Traffic				
Van & truck traffic (6 a.m.-10:59 p.m.)				
Class 2 (low)	1.00***	0.15	6.84	2.72
Class 3 (moderate)	1.31***	0.16	8.21	3.72
Class 4 (high)	1.24***	0.21	6.03	3.45
Class 5 (very high)	2.57***	0.33	7.81	13.01

*** Significant at 99.9%; ** Significant at 99%; * Significant at 95%
[a] Exponentially transformed variables ($e^{-0.001.x}$)
OR: Odds Ratio; SD: Standard Error

Appendix D.5: Logistic model – Model fit and evaluation, diagnostics and inferential tests

	Statistics
Goodness-of-fit	
Log Likelihood	-1063.09
Akaike Information criterion (AIC)	2170.18
Validations of predicted probabilities	
c statistic	0.83
Missclassification rate	0.14
Multicollinearity	
Variance Inflation Factor (max. value)	1.22
Condition Index (max. value)	3.20

	Test statistic	p-value
Overall model evaluations		
Likelihood ratio test (χ^2)	883.35	< 2.2e-16
Wald test (χ^2)	1033.30	0.00
Goodness-of-fit tests		
Hosmer & Lemeshow (χ^2)	14.10	0.08
Le Cessie & Houwelingen (Z)	-1.86	0.06
Spatial autocorrelation tests		
Moran's I for residuals (I)[a]	0.27	< 2.2e-16

[a] Great care is required when analyzing Moran's I since its statistical basis for inference is not well-founded for logistic regression modelling

Appendix D.6: Convergence diagnostics for the autologistic model

Variables	Geweke diagnostic (Z score)[a]			Gelman-Rubin diagnostic[b]		Raftery-Lewis diagnostic (I)[c]			Heidelberg-Welch diagnostic[d]	
	Chain 1	Chain 2	Chain 3	Point estimate	97.5% quantile	Chain 1	Chain 2	Chain 3	Stationarity	Halfwidth
Intercept	-0.48	-1.27	-0.39	1.00	1.00	1.22	1.18	1.30	passed	passed
Autocovariate variable	-1.76	0.71	-1.16	1.00	1.00	1.03	0.97	0.98	passed	passed
Infrastructure										
Complexity index										
Bandwidth = 40 m	-0.15	1.15	-0.20	1.00	1.00	1.03	1.00	1.10	passed	passed
Bridge & no cycle facility	-0.04	1.04	-0.45	1.00	1.00	1.00	1.02	1.00	passed	passed
Contraflow cycling & no crossroad	0.78	2.18	-1.02	1.00	1.00	0.98	0.98	0.99	passed	passed
Cycle facility & crossroad										
Facility 1 (unidir.) & Crossroad 1 (yield/stop)	0.72	0.61	-0.65	1.00	1.00	0.97	0.97	1.03	passed	passed
Facility 2 (bidir.) & Crossroad 1 (yield/stop)	-1.74	-0.75	-0.89	1.00	1.00	1.06	1.00	1.02	passed	passed
Facility 3 (mark.) & Crossroad 3 (traffic light)	0.77	1.11	1.15	1.00	1.00	1.02	1.00	1.00	passed	passed
Facility 3 (mark.) & Crossroad 4 (round.)	-0.45	0.67	0.66	1.00	1.00	1.00	0.98	1.02	passed	passed
Facility 4 (sugg.) & Crossroad 2 (right-of-w.)	-0.34	1.31	-0.66	1.00	1.00	1.09	1.00	1.02	passed	passed
Facility 0 (no facility) & Crossroad 4 (round.)	-0.07	1.25	-0.08	1.00	1.00	1.00	0.97	0.98	passed	passed
Tram tracks										
Class 1 (crossing tram tracks)	-0.95	-1.20	0.22	1.00	1.00	0.97	1.03	1.00	passed	passed
Class 3 (on-road tram tracks)	-0.56	-0.75	-0.45	1.00	1.00	1.07	1.05	1.03	passed	passed
Number of garages (d ≤100m)										
Range 0 (no garage)	1.33	0.33	0.03	1.00	1.00	0.99	1.00	1.00	passed	passed
Distance shopping center	-0.12	1.75	-1.47	1.00	1.00	0.98	0.98	1.00	passed	passed

continued on next page

continued

Variables	Geweke diagnostic (Z score)[a]			Gelman-Rubin diagnostic[b]		Raftery-Lewis diagnostic (I)[c]			Heidelberg-Welch diagnostic[d]	
	Chain 1	Chain 2	Chain 3	Point estimate	97.5% quantile	Chain 1	Chain 2	Chain 3	Stationarity	Halfwidth
Proximity parking-cycle facility										
Parking & Facility 1 (unidirectional)	1.25	-0.91	0.30	1.00	1.00	1.05	1.00	1.03	passed	passed
Parking & Facility 2 (bidirectional)	0.23	-0.20	-0.43	1.00	1.00	1.03	0.95	1.03	passed	passed
Traffic										
Van & truck traffic (6 a.m.-10:59 p.m.)										
Class 2 (low)	1.68	1.20	0.76	1.00	1.00	1.03	1.12	1.05	passed	passed
Class 3 (moderate)	1.50	0.53	-0.43	1.00	1.00	1.05	1.00	1.03	passed	passed
Class 4 (high)	-0.06	-0.44	0.47	1.00	1.00	1.04	1.00	0.98	passed	passed
Class 5 (very high)	-1.04	0.96	0.66	1.00	1.00	1.03	0.95	1.02	passed	passed

[a] Fraction in 1st window = 0.1; fraction in 2nd window = 0.5

[b] Potential scale reduction factors (psrf); multivariate psrf = 1

[c] I = dependence factor; quantile = 0.025; accuracy = +/- 0.005; probability = 0.95

[d] Precision of halfwidth test = 0.1; note that stationarity and halfwidth tests are passed for the 3 Markov chains

Appendix E
Publications and personal contribution to this thesis

Within the framework of the federal research project named SHAPES (funded by Belspo), parts of the two first papers were published in peer-reviewed journals, i.e. in Transport Policy (Chapter 2) and in Transportation Research Part A (Chapter 3). Two other chapters of this thesis (Chapters 4 and 5) will also be submitted in a near future to other peer-reviewed journals (especially those focused on the analysis of road accidents).

On the request of the jury members selected within the framework of this PhD. thesis, I here aim at giving further details about my personal contribution for each of the papers/chapters (see Appendix E.1 for an objective approximate of the % of the time budget I devoted). As first author of these four papers, I conducted most of the research tasks (90-95% of the time budget), especially those related to the review of the literature, the data collection, the analysis of the results and the publication of results. Contribution of co-authors amounts to approximately 5-10% of the total time budget devoted to the papers. In particular, Isabelle Thomas (supervisor of this thesis, UCL-CORE) was of great aid to help me defining the objectives and gave me fruitful comments as regards the data collection, the methodology and the draft version of the papers (especially as regards Chapters 2 and 3). Within the framework of Chapter 3, Claire Dujardin (UCL-CORE) also provided fruitful comments as regards the definition of the variables, the methodological choices and the draft version of the paper. Finally, Bas de Geus (VUB) and Joris Aertsens (VITO) provided some of the data used in Chapter 4.

Also note that, within the framework of the SHAPES project, I contributed as co-author to the publication of other papers and scientific reports (see below), as well as I published two papers in proceedings of international conferences (BIVEC-GIBET 2009, 2011).

Appendix E.1: Approximate % of the time budget devoted to each task and chapter (the remaining % is attributable to co-authors)

	Chap. 2	Chap. 3	Chap. 4	Chap. 5	Mean
Objectives/methodology	60%	70%	95%	95%	**80%**
Literature review	100%	100%	100%	100%	**100%**
Data collection	80%	95%	80%	100%	**89%**
Statistical/geographical treatments	100%	100%	100%	100%	**100%**
Analysis/discussion of the results	90%	90%	95%	95%	**93%**
Writing & revision process	90%	90%	95%	95%	**93%**
Mean	**87%**	**91%**	**94%**	**98%**	**92%**

List of published chapters (peer-reviewed publications)

Vandenbulcke, **G.**, Dujardin, C., Thomas, I., de Geus, B., Degraeuwe, B., Meeusen, R., Int Panis, L. (2011). Cycle commuting in Belgium: Spatial determinants and 're-cycling' strategies. *Transportation Research Part A* 45, 118-137. [http://dx.doi.org/10.1016/j.tra.2010.11.004]

Vandenbulcke, **G.**, Thomas, I., de Geus, B., Degraeuwe, B., Torfs, R., Meeusen, R., Int Panis, L. (2009). Mapping bicycle use and the risk of accidents for commuters who cycle to work in Belgium. *Transport Policy* 16, 77-87. [http://dx.doi.org/10.1016/j.tranpol.2009.03.004]

List of chapters to submit

Vandenbulcke, **G.**, Thomas, I. (et al?). Accident risk when cycling in Brussels: an innovative spatial case-control approach. ***On-going paper***.

Vandenbulcke, **G.**, Thomas, I., de Geus, B., Aertsens, J., Romain, M., Int Panis, L. Reported *versus* unreported cycling accidents: a spatial network analysis for Brussels. ***On-going paper***.

Publications as co-author (not reported in this thesis)

de Geus, B., **Vandenbulcke**, **G.**, Int Panis, L., Thomas, I., Degraeuwe, B., Cumps, E., Aertsens, J., Thomas, I., Torfs, R., Meeusen, R. A prospective

cohort study on minor bicycle accidents: commuter cyclists in Belgium. Accepted for publication in *Accident Analysis and Prevention*.

Aertsens, J., de Geus, B., **Vandenbulcke**, **G.**, Degraeuwe, B., Broekx, S., De Nocker, L., Liekens, I., Mayeres, I., Meeusen, R., Thomas, I., Torfs, R., Willems, H., Int Panis, L. (2010). Commuting by bike in Belgium, the costs of minor accidents. *Accident Analysis and Prevention* 42 (6), 2149-2157.

Int Panis, L., de Geus, B., **Vandenbulcke**, **G.**, Willems, H., Degraeuwe, B, Bleux, N., Mishra, V., Thomas, I., Meeusen, R. (2010). Exposure to particulate matter in traffic: A comparison of cyclists and car passengers. *Atmospheric Environment* 44 (19), 2263-2270.

Bibliography

Aertsens, J., de Geus, B., Vandenbulcke, G., Degraeuwe, B., Broekx, S., De Nocker, L., Liekens, I., Meeusen, R., Thomas, I., Torfs, I., Willems, H., Int Panis, L. (2010). Commuting by bike in Belgium, the costs of minor accidents. *Accident Analysis and Prevention* 42, 2149-2157.

Aguero-Valverde, J., Jovanis, P.P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis and Prevention* 38, 618-625.

Akçakaya, H.R., Atwood, J.L. (1997). A habitat-based metapopulation model of the California Gnatcatcher. *Conservation Biology* 11 (2), 422-434.

Albert de la Bruhèze, A. (1999). *Bicycle use in practice and policy during the twentieth century: similarities and differences of bicycle use in Amsterdam, Eindhoven, Enschede, Southeast-Limburg, Antwerp, Manchester, Copenhagen, Hanover en Basel*. Dutch Ministry of Transport, Public Works and Water Management. [in Dutch]

Alexander, S.M., Paquet, P.C., Logan, T.B., Saher, D.J. (2005). Snow-tracking versus radiotelemetry for predicting wolf-environment relationships in the Rocky Mountains of Canada. *Wildlife Society Bulletin* 33 (4), 1216-1224.

Anderson, T.K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis and Prevention* 41, 359-364.

Angeriz, A., McCombie, J., Roberts, M. (2008). New estimates of returns to scale and spatial spillovers for EU regional manufacturing, 1986-2002. *International Regional Science Review* 31 (1), 62-87.

Anselin, L. (1988). *Spatial econometrics: methods and models*. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Anselin, L. (1992). *Spatial data analysis with GIS: an introduction to application in the social sciences*. Technical report 92-10, National Center for Geographic Information and Analysis, University of California, USA.

Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis* 27, 93-115.

Anselin, L. (1998). Exploratory spatial data analysis in a geocomputational environment. In: Longley, P.A., Brooks, S.M., McDonnell, R., Macmillan, B. (Eds.), *Geocomputation, a Primer*. Wiley, New York, USA.

Anselin, L. (2005). *Exploring spatial data with GeoDa: a workbook*. Spatial Analysis Laboratory and Center for Spatially Integrated Social Science. Available at: [http://geodacenter.asu.edu/learning/tutorials]

Anselin, L. (2007). *Spatial regression analysis in R: a workbook*. Spatial Analysis Laboratory and Center for Spatially Integrated Social Science. Available at: [http://geodacenter.asu.edu/learning/tutorials]

Anselin, L., Bera, A.K., Florax, R., Yoon, M.J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26 (1), 77-104.

Anselin, L., Florax, R. (1995). Small sample properties of tests for spatial dependence in regression models. In: Anselin, L. and Florax, R. (Eds.), *New directions in spatial econometrics*, Springer, Berlin, 21-74.

Anselin, L., Griffith, D.A. (1988). Do spatial effects really matter in regression analysis? *Papers of the Regional Science Association* 65, 11-34.

Anselin, L., Rey, S. (1991). Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23, 112-131.

Augustin, N., Mugglestone, M., Buckland, S. (1996). An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* 33, 339-347.

Aultman-Hall, L., Hall, F. (1998). Ottawa-Carleton commuter cyclist on- and off-road incident rates. *Accident Analysis and Prevention* 30 (1), 29-43.

Aultman-Hall, L., Kaltenecker, M.G. (1999). Toronto bicycle commuter safety rates. *Accident Analysis and Prevention* 31 (6), 675-686.

Bailey, T.C., Gatrell, A.C. (1995). *Interactive Spatial Data Analysis*. Longman, Harlow, UK.

Baller, R.D., Anselin, L., Messner, S.F., Deane, G., Hawkins, D.F. (2001). Structural covariates of U.S. county homicides rates: incorporating spatial effects. *Criminology* 39 (3), 561-590.

Banister, C., Gallent, N. (1998). Trends in commuting in England and Wales becoming less sustainable? *Area* 30 (4), 331-341.

Bergström, A., Magnusson, R. (2003). Potential of transferring car trips to bicycle during winter. *Transportation Research Part A* 37, 649-666.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistic Society B* 36, 192-236.

Besag, J., York, J., Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43 (1), 1-59.

Besag, J., Kooperberg, C. (1995). On conditional and intrinsic autoregression. *Biometrika* 4, 733-746.

Best, N.G., Cowles, M.K., Vines, S.K. (1995). *CODA Manual version 0.30.* MRC Biostatistics Unit, Cambridge, UK. Available online at: [http://www.mrc-bsu.cam.ac.uk/bugs/documentation/coda03/cdaman03.html]

Best, N.G., Richardson, S. (2009). *Introduction to Bayesian analysis using WinBUGS.* Imperial College, Faculty of Medicine, London, UK. Available online at: [http://www.bias-project.org.uk/WBcourse2009/index.htm]

Bivand, R. (2008). Disease Mapping. In: Bivand, R. (Ed.), *Applied spatial data analysis with R*, Springer, New York, USA. Available at: [http://www.springerlink.com/content/978-0-387-78170-9]

Black, W.R., Thomas, I. (1998). Accidents on Belgium's motorways: a network autocorrelation analysis. *Journal of Transport Geography* 6, 23-31.

BMA (British Medical Association) (1992). *Cycling towards health and safety.* Oxford University Press, London, UK.

Bolstad, W.M. (2007). *Introduction to Bayesian Statistics* (2nd Edition). Wiley, New Jersey, USA.

Boots, B.N., Getis, A. (1988). *Point pattern analysis.* Sage Scientific Geography Series 8. Sage Publications, London, UK.

Borgoni, R., Billari, F.C. (2003). Bayesian spatial analysis of demographic survey data: An application to contraceptive use at first sexual intercourse. *Demographic Research* 8 (3), 61-92.

Boufous, S., Williamson, A. (2006). Work-related traffic crashes: a record linkage study. *Accident Analysis and Prevention* 38 (1), 14-21.

Brodie, L., Lyndal, B., Elias, I.G. (2009). Heavy vehicle driver fatalities: learning's from fatal road crash investigations in Victoria. *Accident Analysis and Prevention* 41 (3), 557-564.

Brotons, L., Thuiller, W., Araújo, M.B., Hirzel, A.H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27, 437-448.

BRSI (Belgian Road Safety Institute) (2006). *Accidents de cyclistes en contexte urbain – Trois années (1998-2000) d'accidents corporels de cyclistes sur les voiries régionales de la Région de Bruxelles-Capitale.* IBSR/BIVV, Brussels.

BRSI (Belgian Road Safety Institute, Observatory for road safety) (2008). *Evolution de la sécurité routière en Belgique (2000-2006).* IBSR/BIVV, Brussels.

BRSI (Belgian Road Safety Institute, Observatory for road safety) (2009a). *Rapport thématique Cyclistes: Accidents de la route impliquant des cyclistes (2000-2007).* IBSR/BIVV, Brussels.

BRSI (Belgian Road Safety Institute, Observatory for road safety) (2009b). *Mesure nationale de comportement en matière de vitesse (2003-2007).* IBSR/BIVV, Brussels.

Brunsdon, C. (1995). Estimating probability surfaces for geographical point data: an adaptive kernel algorithm. *Computers and Geosciences* 21, 877-894.

Brunsdon C., A.S. Fotheringham and M. Charlton (1999). Some notes on parametric significance tests for Geographically Weighted Regression. *Journal of Regional Science* 39, 497-524.

Brussels Mobility (2010). *IRIS II – Plan de mobilité (Région de Bruxelles-Capitale).* Brussels Mobility, Brussels. Available at: [http://www.bruxellesmobilite.irisnet.be/articles/la-mobilite-de-demain/en-quelques-mots]

Buehler, R., Pucher, J., Merom, D., Bauman, A. (2011). Active travel in Germany and the U.S. – Contributions of daily walking and cycling to physical activity. *American Journal of Preventive Medicine* 41 (3), 241-250.

Buis, J. (2000). *The economic significance of cycling: a study to illustrate the costs and benefits of cycling policy.* VNG and Interface for Cycling Expertise, De Haag, Netherlands.

Burchell, R.W., Lowenstein, G., Dolphin, W.R., Galley, C.C., Downs, A., Seskin, S., Still, K.G., Moore, T. (2002). *Costs of sprawl – 2000.* TCRP Report 74, Transportation Research Board. National Academy Press, Washington D.C., USA.

Burt, B.A. (2001). Definitions of risk. *Journal of Dental Education* 65 (10), 1007-1008.

Calthorpe, P., Fulton, W. (2001). *The Regional City: Planning for the End of Sprawl*. Island Press: Washington DC (USA).

Cameron, I.C., Harris, N.J., Kehoe, N.J.S. (2001). Tram-related injuries in Sheffield. *Injury* 32 (4), 275-277.

Cervero, R. (1998). *The transit metropolis: A global inquiry.* Island Press, Washington DC, USA.

Cervero, R. (2004). *Developing around transit: strategies and solutions that work.* Urban Land Institute, Washington DC, USA.

Cervero, R., Kockelman, K. (1997). Travel demand and the 3Ds: density, diversity and design. *Transportation Research Part D* 2 (3), 199-219.

Chapman, L. (2007). Transport and climate change: a review. *Journal of Transport Geography* 15, 354-367.

Chefaoui, R.M., Lobo, J.M. (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling* 210, 478-486.

Cho, G., Rodríguez, D.A., Khattak, A.J. (2009). The role of built environment in explaining relationships between perceived and actual pedestrian and bicyclist safety. *Accident Analysis and Prevention* 41, 692-702.

Chong, S., Poulos, R., Olivier, J., Watson, W.L., Grzebieta, R. (2010). Relative injury severity among vulnerable non-motorised road users: Comparative analysis of injury arising from bicycle-motor vehicle and bicycle-pedestrian collisions. *Accident Analysis and Prevention* 42, 290-296.

Clark, J.S. (2005). Why environmental scientists are becoming Bayesians. *Ecology Letters* 8, 2-14.

City of Paris (2008). *Le bilan des déplacements en 2008 à Paris.* L'observatoire des déplacements à Paris, Paris. Available at: [http://www.paris.fr/portail/deplacements/Portal.lut?page_id=9071&document_type_id=5&document_id=64255&portlet_id=21968]

Cohen, J., Cohen, P., West, S.G., Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd Edition). Lawrence Erlbaum Associates, New Jersey, USA.

Congdon, P. (2003). *Applied Bayesian modelling.* Wiley, Chichester, UK.

Conger, A.J. (1974). A revised definition for suppressor variables: a guide to their identification and interpretation. *Educational and Psychological Measurement* 34, 35-46.

Corrado, L., Fingleton, B. (2011). *Multilevel Modelling with Spatial Effects.* Discussion Papers in Economics, University of Strathclyde, Glasgow, UK.

Cressie, N.A.C. (1993). *Statistics for spatial data.* Wiley, New York, USA.

Curtis, C., Headicar, P. (1997). Targeting travel awareness campaigns – which individuals are more likely to switch from car to other transport for the journey to work? *Transport Policy* 4 (1), 57-65.

Dai, D., Taquechel, E., Steward, J., Strasser, S. (2010). The impact of built environment on pedestrian crashes and the identification of crash clusters on an urban University campus. *Western Journal of Emergency Medicine* XI (3), 294-301.

Daniel, S., Geurts, K. (2004). *Verkeersonveiligheid in de provincie Limburg.* Rapport SN-2004-04. Steunpunt Verkeersveiligheid, Diepenbeek, Belgium.

Daniels, S., Nuyts, E., Wets, G. (2008). The effects of roundabouts on traffic safety for bicyclists: an observational study. *Accident Analysis and Prevention* 40, 518-526.

Daniels, S., Brijs, T., Nuyts, E., Wets, G. (2009). Injury crashes with bicyclists at roundabouts: influence of some location characteristics and the design of cycle facilities. *Accident Analysis and Prevention* 40, 141-148.

Darlington, R.B. (1978). Multiple regression in psychological research and practice. *Psychological Bulletin* 69, 161-182.

Daya, S. (2000). Understanding statistics: odds ratio. Evidence-based Obstetrics and *Gynecology* 2, 84-85.

Deboosere, P., Demarest, S., Lorant, V., Miermans, P-J., Portet, M-I., Van Oyen, H. (2006). *General socio-economic survey 2001 – Monography n°1: Health and informal care.* FPS Economy, SMEs, Self-employed and Energy, Brussels, 175 p. [in French/Dutch]

Deboosere, P., Lesthaeghe, R., Surkyn, J., Willaert, D., Boulanger, P-M., Lambert, A., Lohlé-Tart, L. (2009). *General socio-economic survey 2001 – Monography n°4: Households and families in Belgium.* FPS Economy, SMEs, Self-employed and Energy, Brussels, 170 p. [in French/Dutch]

De Borger, B., Peirson, J., Vickerman, R.W. (2001). An overview of policy instruments. In: De Borger, B. and Proost, S. (Eds.), *Reforming Transport Pricing in the European Union*, Edward Elgar, Cheltenham, 37-50.

De Bourdeaudhuij, I., Teixeira, P.J., Cardon, G., Deforche, B. (2005). Environmental and psychosocial correlates of physical activity in Portuguese and Belgium adults. *Public Health Nutrition* 8 (7), 886-895.

Deckers, B., Verheyen, K., Hermy, M., Muys, B. (2005). Effects of landscape structure on the invasive spread of black cherry *Prunus serotina* in an agricultural landscape in Flanders, Belgium. *Ecography* 28, 99-109.

de Geus, B. (2007). *Cycling to work: psychosocial and environmental factors associated with cycling and the effect of cycling on fitness and health indexes in an untrained working population.* Ph.D. Thesis in Physical Education, Vrije Universiteit Brussel (VUB), Brussels, Belgium.

de Geus, B., Van Hoof, E., Aerts, I., Meeusen, R. (2008a). Cycling to work: influence on indexes of health in untrained men and women in Flanders. Coronary heart disease and quality of life. *Scandinavian Journal of Medicine and Science in Sports* 18 (4), 498-510.

de Geus, B., De Bourdeaudhuij, I., Jannes, C., Meeusen, R. (2008b). Psychosocial and environmental factors associated with cycling for transport among a working population. *Health Education Research* 23 (4), 697-708.

de Geus, B., Joncheere, J., Meeusen, R. (2009). Commuter cycling: effect on physical performance in untrained men and women in Flanders. Minimum dose to improve indexes of fitness. *Scandinavian Journal of Medicine and Science in Sports* 19 (2), 179-187.

de Geus, B., Vandenbulcke, G., Int Panis, L., Degraeuwe, B., Cumps, E., Aertsens, J., Thomas, I., Torfs, R., Meeusen, R. A prospective cohort study on minor accidents involving commuter cyclists in Belgium. Accepted for publication in *Accident Analysis and Prevention*.

de Hartog, J.J., Boogaard, H., Nijland, H., Hoek, G. (2010). Do the health benefits of cycling outweigh the risks? *Environmental Health Perspectives* 118 (8), 1109-1116. doi:10.1289/ehp.0901747.

de Lapparent, M. (2005). Individual cyclists' probability distributions of severe/fatal crashes in large french urban areas. *Accident Analysis and Prevention* 37, 1086-1092.

Delmelle, E.C., Thill, J-C. (2008). Urban bicyclists – a spatial analysis of adult and youth traffic hazard intensity. *Transportation Research Record*.

De Mol, J., Lammar, P. (2006). Helft verkeersslachtoffers komt niet in statistieken. *Verkeersspecialist* 130, 15-18.

De Witte, A., Macharis, C. (2010). Commuting to Brussels: how attractive is 'free' public transport? *Brussels Studies* 37, 19 April 2010. Available online at:
[http://www.brusselsstudies.be/PDF/EN_124_BruS37EN.pdf]

Diggle, P.J. (1983). *Statistical analysis of spatial point patterns*. Academic Press, London, UK.

DGSEI (Directorate-General Statistics and Economic Information) (2000). Downloadable files: Society, mobility and transport. Available at:
[http://statbel.fgov.be/figures/download_fr.asp#mob]

DGSEI (Directorate-General Statistics and Economic Information) (2001a). *National household survey on mobility. Realization and results*. Research contract MD/13/036, FUNDP-GRT-LV-UIA-IW-INS. Belgian Federal Science Policy Office, Brussels, Belgium. Available (in French) at: [http://www.belspo.be/belspo/fedra/proj.asp?l=en&COD=MD/DD/18]

DGSEI (Directorate-General Statistics and Economic Information) (2001b). Downloadable files. Available at: [http://statbel.fgov.be/figures/download_fr.asp]

DGSEI (Directorate-General Statistics and Economic Information) (2004). Downloadable files: Territory and environment. Available at: [http://statbel.fgov.be/figures/download_fr.asp#1]

DGSEI (Directorate-General Statistics and Economic Information) (2002-2005). Downloadable files: Society, road accidents. Available at: [http://www.statbel.fgov.be/figures/download_fr.asp#acc]

Dickinson, J.E., Kingham, S., Copsey, S., Pearlman Hougie, D.J. (2003). Employer travel plans, cycling and gender: will travel plan measures improve the outlook for cycling to work in the UK? *Transportation Research Part D* 8, 53-67.

Diez-Roux, A.V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health* 21, 171-192.

Diggle, P.J. (1990). A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society A* 153 (3), 349-362.

Dill, J., Voros, K. (2007). *Factors affecting bicycling demand: initial survey findings from the Portland Region*. Transportation Research Board, Washington D.C, USA.

Dobruszkes, F., Marissal, P. (1994). La problématique des déplacements à Bruxelles. *Revue Belge de Géographie* 118 (3), 133-153.

Doom, R., Derweduwen, P. (2005). *Optimalisatie van de verkeersongevallenstatistieken*. Research contracts CP/02/392, CP/F1/391 (final report) for BRSI, CDO & Belgian Science Policy, Belgium.

Dormann, C.F. (2007). Assessing the validity of autologistic regression. *Ecological Modelling* 207, 234-242.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30, 609-628.

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. *Journal of Educational and Behavioral Statistics* 20, 115-147.

Dujardin, C., Thomas, I., Tulkens, H. (2007). Quelles frontières pour Bruxelles ? Une mise à jour. *Reflets et Perspectives de la Vie Economique* 46, 156-176.

Dumbaugh, E., Rae, R. (2009). Safe urban form: revisiting the relationship between community design and traffic safety. *Journal of the American Planning Association* 75 (3), 309-329.

Duncan, C., Jones, K., Moon, G. (1998). Context, composition and heterogeneity: using multilevel models in health research. *Social Science and Medicine* 46, 97-117.

Dutch Ministry of Transport, Public Works and Water Management (2007). Cycling in the Netherlands. Available online at: [http://www.fietsberaad.nl/index.cfm?lang=en&repository=Cycling+in+the+Netherlands]

Ebdon, D. (1985). *Statistics in Geography* (2nd edition). Basil Blackwell, Oxford, UK.

EC (European Commission) (2000). *Villes cyclables, villes d'avenir*. Office des publications officielles des Communautés européennes, Luxembourg. Available at: [http://ec.europa.eu/environment/archives/cycling/cycling_fr.pdf]

EEA (European Environment Agency) (2007). *Air Pollution in Europe 1990-2004*. Office for Official Publications of the European Communities, Luxembourg. Available at: [http://www.eea.europa.eu/publications/eea_report_2007_2]

Erdogan, S., Ylmaz, I., Baybura, T., Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. *Accident Analysis and Prevention* 40 (1), 174-181.

Eksler, V. (2008). Exploring spatial structure behind the road mortality of regions in Europe. *Applied Spatial Analysis* 1 (2), 133-150.

Eksler, V., Lassarre, S. (2008). Evolution of road risk disparities at small-scale level: Example of Belgium. *Journal of Safety Research* 39, 417-427.

Eluru, N., Bhat, C.R., Hensher, D.A. (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention* 40, 1033-1054.

Elvik, R. (2006). Laws of accident causation. *Accident Analysis and Prevention* 38, 742-747.

Elvik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis and Prevention* 41 (4), 849-855.

Emmerson, P., Riley, T.J., Davies, D.G. (1998). The impact of weather on cycle flows. *Traffic Engineering and Control* 39 (4), 238-243.

Engler, R., Guisan, A., Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* 41, 263-274.

EROS (Earth Resources Observation and Science Center) (2002). Shuttle radar topography mission (SRTM) – Elevation data set (Belgium). Available at: [http://eros.usgs.gov/]

ERSO (European Road Safety Observatory) (2006). *Pedestrians & Cyclists*. Retrieved 4-17-2011 from www.erso.eu.

Ervin, G.N. (2009). Using GAP data in invasive plan ecology and management. *Gap Analysis Bulletin* 16, 34-41.

ESRI (Environmental Systems Research Institute, Inc.) (2009). *Measuring Geographic Distributions toolset*. Redlands, CA (USA).

ESRI (Environmental Systems Research Institute, Inc.) (2010). *Geocoding and address management*. Redlands, CA (USA).

EU (European Union) (2003). *Energy and transport in figures*. European Commission, Directorate-General for Energy and Transport, Brussels. Available online at: [http://gasunie.eldoc.ub.rug.nl/root/2004/2999175/]

Federal Police (2000-2002). *Bicycle Theft during the Period 2000-2006, on the Zonal and Communal Scales*. General Commissioner's Office, Directorate of Operational Police Information, Brussels, Belgium (available in French and Dutch).

Federal Police (2006). *Moniteur de sécurité 2006*. Direction de la banque de données nationale, Bruxelles. Available from: [http://www.polfed-fedpol.be/pub/veiligheidsMonitor/2006/monitor2006_fr.php]

Fernandez, G. (2002). *Data mining using SAS applications*. Chapman & Hall/CRC, London, UK.

Ferrier, S., Watson, G., Pearce, J., Drielsma, M. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation* 11, 2275-2307.

Fisher, M.M., Bartkowska, M., Riedl, A., Sardadvar, S., Kunnert, A. (2009). The impact of human capital on regional labor productivity in Europe. *Letters in Spatial and Resource Sciences.* Available at: [http://www.springerlink.com/content/l04140hw51p12k77/]

Flahaut, B. (2004). Impact of infrastructure and local environment on road unsafety: Logistic modelling with spatial autocorrelation. *Accident Analysis and Prevention* 36 (6), 1055-1066.

Flahaut, B., Mouchart, M., San Martin, E., Thomas, I. (2003). The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. *Accident Analysis and Prevention* 35 (6), 991-1004.

Forester, J. (1994). *Bicycle transportation – A handbook for cycling transportation engineers.* MIT, Cambridge, USA.

Fotheringham, A.S., Brunsdon, C., Charlton, M. (2000). *Quantitative geography: perspectives on spatial data analysis.* Sage Publications, London, UK.

Fotheringham, A.S., Densham, P., Curtis, A. (1995). The zone definition problem in allocation modelling in location-allocation modelling. *Geographical Analysis* 27, 60-77.

Fotheringham, A.S., Wong, D.W.S. (1991). The modifiable unit problem in multivariate statistical analysis. *Environment and Planning A* 23, 1025-1044.

French, K.M., Jones, K. (2006). Impact of definition on the study of avoidable mortality: geographical trends in British deaths 1981-1998 using Charlton and Holland's definitions. *Social Science and Medicine* 62, 1443-1456.

Frey, H. (1999). *Designing the city: towards a more sustainable urban form.* Taylor & Francis, London, UK.

Friedman, L., Wall, M. (2005). Graphical view of suppression and multicollinearity in multiple linear regression. *The American Statistician* 59 (2), 127-136.

Gatersleben, B., Appleton, K.M. (2007). Contemplating cycling to work: attitudes and perceptions in different stages of change. *Transportation Research Part A* 41, 302-312.

Gatrell, A.C., Bailey, T.C., Diggle, P.J., Rowlingson, B.S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers* 21 (1), 256-274.

Gauderman, W.J., Vora, H., McConnell, R., Berhane, K., Gilliland, F., Thomas, D., Lurmann, F., Avol, E., Kunzli, N., Jerrett, M., Peters, J. (2007). Effect of exposure to traffic on lung development from 10 to 18 years of age: a cohort study. *The Lancet* 369, 571-577.

Geertman, S.C.M., Ritsema van Eck, J.R. (1995). GIS and models of accessibility potential: an application in planning. *International Journal of Geographical Information Systems* 9 (1), 67-80.

Gelfand, A., Gosh, S. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika* 85, 1-13.

Gelman, A. (2006). Multilevel (hierarchical) modeling: what it can do and cannot do. *Technometrics* 48, 432-435.

Gelman, A., Carlin, J., Stern, H., Rubin, D. (1995). *Bayesian data analysis*. Chapman & Hall, London, UK.

Gelman, A., Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, New York, USA.

Gelman, A., Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457-511.

Gelman, A., Shor, B., Bafumi, J., Park, D. (2007). Rich state, poor state, red state, blue state: what's the matter with Connecticut? *Quarterly Journal of Political Science* 2, 345-367.

Getis, A., Ord, J. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24, 189-206.

Geurs, K.T., Ritsema van Eck, J.R. (2001). *Accessibility measures: review and applications*. RIVM Report 408505 006. National Institute of Public Health and the Environment, Bilthoven, the Netherlands. Available from: <http://www.rivm.nl/bibliotheek/rapporten/408505006.html>

Geurs, K.T., van Wee, B. (2004). Accessibility of land-use and transport strategies: review and research directions. *Journal of Transport Geography* 12, 127-140.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.). *Bayesian Statistics* (Vol. 4), pages 169-194, Claredon Press, Oxford, UK.

Goldstein, H. (1999). *Multilevel Statistical Models*. 1^{st} Internet Edition. Edward Arnold, London, UK.

Graw, M., König, H.G. (2002). Fatal pedestrian-bicycle collisions. *Forensic Science International* 126, 241-247.

Greater Lyon (2009). *Comptage des vélos – En trois ans, deux fois plus de vélos.* Grand Lyon, Communauté urbaine. Available at: [http://www.grandlyon.com/Comptage-des-velos.2231.0.html]

Greenland, S. (2000). Principles of multilevel modelling. *International Journal of Epidemiology* 29, 158-167.

Greibe, P. (2003). Accident prediction models for urban roads. *Accident Analysis and Prevention* 35, 273-285.

Grimes, D.A., Schulz, K.F. (2005). Compared to what? Finding controls for case-control studies. *The Lancet* 365, 1429-1433.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S., Peterson, A.T. (2007). What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monography* 77 (4), 615-630.

Haining, R., Law, J., Griffith, D. (2009). Modelling small area counts in the presence of overdispersion and spatial autocorrelation. *Computational Statistics and Data Analysis* 53, 2923-2937.

Hansen, W.G. (1959). How accessibility shapes land use. *Journal of American Institute of Planners* 25 (1), 73-76.

Haque, M.M., Chin H.C., Huang, H. (2010). Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. *Accident Analysis and Prevention* 42, 203-212.

Harris, M.A., Reynolds, C.C.O., Winters, M., Chipman, M., Cripton, P.A., Cusimano, M.D., Teschke, K. (2011). The bicyclists' injuries and the cycling environment study: a protocol to tackle methodological issues facing studies of bicycling safety. *Injury Prevention* (in press).

Hauer, E., Hakkert, A.S. (1988). Extent and some implications of incomplete accident reporting. *Transportation Research Record* 1185, 1-10.

Haynes, R., Lovett, A., Sünnenberg, G. (2003). Potential accessibility, travel time, and consumer choice: geographical variations in general medical practice registrations in Eastern England. *Environment and Planning A* 35, 1733-1750.

Hearst, M.O., Oakes, J.M., Johnson, P.J. (2008). The effect of racial residential segregation on black infant mortality. *American Journal of Epidemiology* 168, 1247-1254.

Hedelin, A., Björnstig, U., Brismar, B. (1996). Trams – A risk factor for pedestrians. *Accident Analysis and Prevention* 28 (6), 733-738.

Hels, T., Orozova-Bekkevold, I. (2007). The effect of roundabout design features on cyclist accident rate. *Accident Analysis and Prevention* 39, 300-307.

Heidelberg, P., Welch, P. (1992). Simulation run length control in the presence of an initial transient. *Operations Research* 31, 1109-1144.

Heinen, E., van Wee, B., Maat, K. (2009). Impact of work-related factors on levels of bicycle commuting. In: *88th Meeting of the Transportation Research Board*, 11-15 January 2009, Paper 09-1212, Washington DC, USA.

Heinen, E., van Wee, B., Maat, K. (2010). Commuting by bicycle: An overview of the literature. *Transport Reviews* 30 (1), 59-96.

Hels, T., Orozova-Bekkevold, I. (2007). The effect of roundabout design features on cyclist accident rate. *Accident Analysis and Prevention* 39, 300-307.

Hengl, T., Sierdsema, H., Radović, A., Dilo, A. (2009). Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. *Ecological Modelling* 220, 3499-3511.

Hertel, O., Hvidberg, M., Ketzel, M., Storm, L., Stausgaard, L. (2008). A proper choice of route significantly reduces air pollution exposure – a study on bicycle and bus trips in urban streets. *Science of the Total Environment* 389, 58-70.

Hoeting, J.A., Leecaster, M., Bowden, D. (2000). An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics* 5 (1), 102-114.

Hopkinson, P., Wardman, M. (1996). Evaluating the demand for new cycle facilities. *Transport Policy* 3 (4), 241-249.

Horst, P. (1941). The role of predictor variables which are independent of the criterion. *Social Science Research Bulletin* 48, 431-436.

Hossain, M.M., Lawson, A.B. (2009). Approximate methods in Bayesian point process spatial models. *Computational Statistics and Data Analysis* 53, 2831-2842.

Hox, J. (2002). *Multilevel Analysis: Techniques and Applications.* Lawrence Erlbaum Associates, Mahwah, NJ, USA.

Hubert, J-P., Toint, P. (2002). La mobilité quotidienne des belges. Presses universitaires de Namur, Namur.

Hunt, J.D., Abraham, J.E. (2007). Influences on bicycle use. *Transportation* 34, 453-470.

Iacono, M., Krizek, K., El-Geneidy, A. (2010). Measuring non-motorized accessibility: issues, alternatives, and execution. *Journal of Transport Geography* 18, 133-140.

Int Panis, L., de Geus, B., Vandenbulcke, G., Willems, H., Degraeuwe, B., Bleux, N., Mishra, V., Thomas, I., Meeusen, R. (2010). Exposure to particulate matter in traffic: a comparison of cyclists and car passengers. *Atmospheric Environment* 44, 2263-2270.

Int Panis, L., Meeusen, R., Thomas, I., Aertsens, J., de Geus, B., Degraeuwe, B., Frère, J., Torfs, R., Vandenbulcke, G., Willems, H., Jacobs, L., Nawrot, T. (2011). *SHAPES – Systematic analysis of health risks and physical activity associated with cycling policies.* Research contract SD/HE/03A (final report) for Belgian Science Policy, Belgium.

Ishihama, F., Takeda, T., Oguma, H., Takenaka, A. (2010). Comparison of effects of spatial autocorrelation on distribution predictions of four rare plant species in the Watarase wetland. *Ecological Research* 25, 1057-1069.

IRCEL-CELINE (Belgian Interregional Cell for the Environment) (2000-2005). Measurements on particulate matter concentrations (PM10). Available at: [http://www.irceline.be/]

Jacobsen, P.L. (2003). Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury Prevention* 9, 205-209.

Jensen, M. (1999). Passion and heart in transport – a sociological analysis on transport behaviour. *Transport Policy* 6, 19-33.

Joanne-Peng, C-Y., Lida Lee, K., Ingersoll, G.M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research* 96 (1), 3-14.

Johnson, P.J., Oakes, J.M., Anderton, D.L. (2008). Neighborhood poverty and American Indian infant death: are the effects identificable? *Annals of Epidemiology* 18, 552-559.

Johnston, R., Jones, K., Propper, C., Burgess, S. (2007). Region, local context, and voting at the 1997 general election in England. *American Journal of Political Science* 51, 640-654.

Jones, A.P., Langford, I.H., Bentham, G. (1996). The application of *K*-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England. *Social Science and Medicine* 42 (6), 879-885.

Kaplan, J. (1976). *Characteristics of the regular adult bicycle user.* Federal Highway Administration: Washington DC, USA.

Kelejian, H.H. and D.P. Robinson (1992). Spatial autocorrelation: a new computationally simple test with an application to per capita county police expenditures. *Regional Science and Urban Economics* 22, 317-31.

Kelsall, J.E., Wakefield, J.C. (1999). Discussion of "Bayesian models for spatially correlated disease and exposure data" by Best, N., Waller, I., Thomas, A., Conlon, E., Arnold, R. In: Bernardo, J.M., Berger, J.O., David, A.P., Smith, A.F.M. Oxford University Press, Oxford, UK.

Kéry, M. (2010). *Introduction to WinBUGS for ecologists. A Bayesian approach to regression, ANOVA, mixed models and related analyses.* Academic Press, Amsterdam, the Netherlands.

Khan, G., Santiago-Chaparro, K.R., Qin, X., Noyce, D.A. (2009). Application and integration of lattice data analysis, network K-functions, and GIS to study ice-related crashes. *Transportation Research Record: Journal of the Transportation Research Board* 2136, 67-76.

Kim, J-K., Kim, S., Ulfarsson, G.F., Porrello, L.A. (2007). Bicyclist injury severities in bicycle-motor vehicle accidents. *Accident Analysis and Prevention* 39, 238-251.

Kingham, S., Dickinson, J., Copsey, S. (2001). Travelling to work: will people move out of their cars. *Transport Policy* 8, 151-160.

Kingham, S., Sabel, C.E., Bartie, P. (2011). The impact of the 'school run' on road traffic accidents: a spatio-temporal analysis. *Journal of Transport Geography* 19, 705-711.

Kirby, D.K., LeSage, J.P. (2009). Changes in commuting to work times over the 1990 to 2000 period. *Regional Science and Urban Economics* 39 (4), 460-471.

Kitamura, R., Mokhtarian, P., Laidet, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation* 24, 125-158.

Klop, J.R., Khattak, A.J. (1999). Factors influencing bicycle crash severity on two-lane, undivided roadways in North-Carolina. *Transportation Research Record* 1674, 78-85.

Knowles, R.D. (2006). Transport shaping space: differential collapse in time-space. *Journal of Transport Geography* 14, 407-425.

Koetse, M.J., Rietveld, P. (2009). The impact of climate change and weather on transport: an overview of empirical findings. *Transportation Research Part D* 14 (3), 205-221.

Koop, G. (2003). *Bayesian econometrics*. Wiley Publishers, West Sussex, UK.

Krizek, K.J., Forsyth, A., Baum, L. (2010). *Walking and cycling international literature review. Final report.* Department of Transport, Melbourne, Australia. Available at: [http://www.transport.vic.gov.au/DOI/DOIElect.nsf/$UNIDS+for+Web+Display/70D43560D1141DDFCA2575E8000BA1EE/$FILE/WalkingCyclingLiteratureReview.pdf]

Kumara, S.S.P., Chin, H.C. (2005). Application of Poisson underreporting model to examine crash frequencies at signalized three-legged intersections. *Transportation Research Record* 1908, 46-50.

Lammar, P., Hens, L. (2004). *Onderzoek naar het gebruik van ziekenhuisgegevens: Minimale Klinische Gegevens.* Steunpuntverkeersveiligheid, Diepenbeek, Belgium.

Lammar, P., Hens, L. (2006). *Haalbaarheidsstudie voor de corrective van de ongevallengegevens – Eindrapport.* Steunpuntverkeersveiligheid, Diepenbeek, Belgium.

Larsen, J., El-Geneidy, A. (2010). *Build it, but where? The use of Geographic Information Systems in identifying optimal location for new cycling infrastructure.* Paper presented at the 89th Transportation Research Board Annual Meeting. Washington DC, USA.

Law, J., Haining, R. (2004). A Bayesian approach to modeling binary data: the case of high-intensity crime areas. *Geographical Analysis* 36 (3), 197-216.

Law, J., Haining, R., Maheswaran, R., Pearson, T. (2006). Analyzing the relationship between smoking and coronary heart disease at the small area level: A Bayesian approach to spatial modeling. *Geographical Analysis* 38, 140-159.

Lawson, A.B., Browne, W.J., Vidal Rodeiro, C.L. (2003). *Disease mapping with WinBUGS and MLwiN.* Wiley & Sons, West Sussex, UK.

Lawson, A.B. (2009). *Bayesian disease mapping. Hierarchical modelling in spatial epidemiology.* Chapman and Hall, London, UK.

Le Gallo, J. (2004). Hétérogénéité spatiale. Principes et méthodes. *Economie et prévision* 1 (162), 151-172.

Le Gallo, J., Ertur, C. (2003). Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1985. *Papers in Regional Science* 82, 175-201.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology* 74, 1659-1673.

Legendre, P., Dale, M.R.T., Fortin, M-J., Gurevitch, J., Hohn, M., Myers, D. (2002). The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25, 601-615.

LeSage, J.P., Fisher, M.M. (2008). Spatial growth regressions: Model specification, estimation and interpretation. *Spatial Economic Analysis* 3 (3), 275-304.

LeSage, J.P., Pace, R.K. (2009). *Introduction to spatial econometrics*. Taylor & Francis CRC Press, Boca Raton, USA.

Levine, N., Kim, K.E., Nitz, L.H. (1995a). Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. *Accident Analysis and Prevention* 27 (5), 663-674.

Levine, N., Kim, K.E., Nitz, L.H. (1995b). Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis and Prevention* 27 (5), 675-685.

Li, L., Zhu, L., Sui, D.Z. (2007). A GIS-based Bayesian approach for analyzing spatial-temporal patterns of intra-city motor vehicle crashes. *Journal of Transport Geography* 15 (4), 274-285.

Litman, T. (1994). Quantifying bicycling benefits for achieving TDM objectives. *Transportation Research Record* 1441, 134-140.

Litman, T. (1995). Land use impact costs of transportation. *World Transport Policy and Practice* 1 (4), 9-16.

Litman, T. (2004). *Quantifying the benefits of nonmotorized transportation for achieving mobility management objectives*. Victoria Transport Policy Institute, Canada.

Liu, Y., Jarrett, D. (2008). Spatial distribution of road crashes in Great Britain. *Journal of Maps*, 91-107.

Long, J.S., Ervin, L.H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician* 54 (3), 217–224.

Loo, B.P.Y., Tsui, K.L. (2010). Bicycle crash casualties in a highly motorized city. *Accident Analysis and Prevention* 42 (6), 1902-1907.

Lord, D., Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A* 44, 291-305.

Lord, D., Washington, S.P., Ivan, J.N. (2005). Poisson, Poisson-Gamma and Zero-Inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. *Accident Analysis and Prevention* 37 (1), 35-46.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D. (2000). WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* 10, 325-337.

Lusk, A.C., Furth, P.G., Morency, P., Miranda-Moreno, L.F., Willett, W.C., Dennerlein, J.T. (2011). Risk of injury for bicycling on cycle tracks versus in the street. *Injury Prevention*. [doi: 10.1136/ip.2010.028696]

Ma, J. (2009). *Bayesian analysis of underreporting Poisson regression model with an application to traffic crashes on two-lane highways*. Paper #09-3192. Presented at the 88th Annual Meeting of the Transportation Research Board, Washington DC, USA.

Martens, K. (2004). The bicycle as a feedering mode: experiences from three European countries. *Transportation Research Part D* 9, 281-294.

Martens, K. (2007). Promoting bike-and-ride: The Dutch experience. *Transportation Research Part A* 41, 326-338.

Mayou, R., Bryant, B. (2003). Consequences of road traffic accidents for different types of road user. *Injury* 34, 197-202.

McCarthy, M., Gilbert, K. (1996). Cyclist road deaths in London 1985-1992: drivers, vehicles, manoeuvres and injuries. *Accident Analysis and Prevention* 28 (2), 275-279.

McClintock, H., Cleary, J. (1996). Cycle facilities and cyclists' safety – experience from Greater Nottingham and lessons for future cycling provision. *Transport Policy* 3, 67-77.

McNemar, Q. (1969). *Psychological statistics*. Wiley, New York, USA.

Mérenne-Schoumaker, B.,Van der Haegen, H., Van Hecke, E., 1999. *General socio-economic survey of the Belgian population and for the housing (1991): Work and school migrations*. Directorate-General Statistics and Economic Information (DGSEI) and Belgian Federal Science Policy Office, Brussels, Belgium. [in French]

Meurs, H., Haaijer, R. (2001). Spatial structure and mobility. *Transportation Research Part D* 6, 429-446.

MF (Mobile Flanders), 2002. Flemish Cycling Plan. Department Mobility and Public Transport. Available online at: <http://www.mobielvlaanderen.be/vtf/vtf01.php?a=14> [in Dutch]

Miaou, S-P., Song, J.J., Mallick, B.K. (2003). Roadway traffic crash mapping: A space-time modeling approach. *Journal of Transportation and Statistics* 6 (1), 33-57.

Miller, J., Franklin, J., Aspinall, R. (2007). Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling* 202, 225-242.

Mohan, J., Twigg, L., Barnard, S., Jones, K. (2005). Social capital, geography and health: a small-area analysis for England. *Social Science and Medicine* 60, 1267-1283.

Møller, M., Hels, T. (2008). Cyclists' perception of risk in roundabouts. *Accident Analysis and Prevention* 40, 1055-1062.

Moran, P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society B* 10, 243-251.

Moudon, A.V., Lee, C., Cheadle, A.D., Collier, C.W., Johnson, D., Schmid, T.L., Weather, R.D. (2005). Cycling and the built environment, a US perspective. *Transportation Research Part D* 10, 245-261.

Myint, S.W. (2008). An exploration of spatial dispersion, pattern and association of socio-economic functional units in an urban system. *Applied Geography* 28, 168-188.

Nankervis, M. (1999). The effect of weather and climate on bicycle commuting. *Transportation Research Part A* 33, 417-431.

Nicholson, A.J. (1985). The variability of accident counts. *Accident Analysis and Prevention* 17 (1), 47-56.

Nielsen, O.H. (1990). Safe routes to school in Odense, Denmark. In Tolley, R. (Ed.), *The greening of urban transport: planning for walking and cycling in Western cities*, Behalven Press, London, UK.

Noël, N. (2003). *Formes urbaines, aménagements routiers et usage de la bicyclette*. Ph.D. Thesis, Université Laval, Québec, Canada.

Noland, R.B., Kunreuther, H. (1995). Short-run and long-run policies for increasingbicycle transportation for daily commuter trips. *Transport Policy* 2 (1), 67-79.

Noland, R.B., Quddus, M.M. (2004). A spatially disaggregate analysis of road casualties in England. *Accident Analysis and Prevention* 36, 973-984.

Ntzoufras, I. (2009). *Bayesian modelling using WinBUGS*. John Wiley & Sons, New Jersey, USA.

Oakes, J.M. (2004). Causal inference and the relevance of social epidemiology. *Social Science and Medicine* 58 (10), 1969-1971.

Oakes, J.M. (2006). Propensity score matching for social epidemiology. In Oakes, J.M., Kaufman, J. (Eds.), *Methods in Social Epidemiology*, pages 370-392, Jossey-Bass Editions.

Oakes, J.M. (2009). Commentary: Individual, ecological and multilevel fallacies. *International Journal of Epidemiology* 38 (2), 361-368.

OECD (Organisation for Economic Co-operation and Development) (2006). *Speed management*. Available online at: [http://www.internationaltransportforum.org/Pub/pdf/06Speed.pdf]

Oja, P., Titze, S., Bauman, A., de Geus, B., Krenn, P., Reger-Nash, B., Kohlberger, T. (2011). Health benefits of cycling: a systematic review. *Scandinavian Journal of Medicine & Science in Sports* 21. [doi: 10.1111/j.1600-0838.2011.01299.x]

Okabe, A., Okunuki, K-I., Shiode, S. (2006a). SANET: a toolbox for spatial analysis on a network. *Geographical Analysis* 38, 57-66.

Okabe, A., Okunuki, K-I., Shiode, S. (2006b). The SANET toolbox: new methods for network spatial analysis. *Transactions in GIS* 10 (4), 535-550.

Okabe, A., Satoh, T. (2006). Uniform network transformation for points pattern analysis on a non-uniform network. *Journal of Geographical Systems* 8 (1), 25-37.

Okabe, A., Satoh, T., Sugihara, K. (2009). A kernel density estimation method for networks, its computational method and a GIS-based tool. *International Journal of Geographical Information Science* 23 (1), 7-32.

Okabe, A., Yamada, I. (2001). The *K*-function method on a network and its computational implementation. *Geographical Analysis* 33 (3), 271-290.

Olivier, F., Wotherspoon, S.J. (2006). Modelling habitat selection using presence-only data: Case study of a colonial hollow nesting bird, the snow petrel. *Ecological Modelling* 195, 187-204.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and Planning A* 16, 17-31.

Openshaw, S., Taylor, P.J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In Wrigley, N., Bennet, R.J. (Eds.), *Statistical applications in the spatial sciences*, pages 127-144, Pion Editions.

Ortúzar, J. de D., Iacobelli, A., Valeze, C. (2000). Estimating demand for a cycle-way network. *Transportation Research Part A* 34, 353-373.

O'Sullivan, D., Unwin, D. (2002). *Geographic Information Analysis*. Wiley & Sons, Hoboken, New Jersey, USA.

Pai, C-W. (2011). Overtaking, rear-end, and door crashes involving bicycles: An empirical investigation. *Accident Analysis and Prevention* 43, 1228-1235.

Parkin, J., Wardman, M., Page, M. (2007). Models of perceived cycling risk and route acceptability. *Accident Analysis and Prevention* 39 (2), 364-371.

Parkin, J., Wardman, M., Page, M. (2008). Estimation of the determinants of bicycle mode share for the journey to work using census data. *Transportation* 35 (1), 93-109.

Parkin, J., Meyers, C. (2010). The effect of cycle lanes on the proximity between motor traffic and cycle traffic. *Accident Analysis and Prevention* 42, 159-165.

PBIC (Pedestrian and Bicycle Information Center) (2007). *Introduction to safe routes to school: the health, safety and transportation nexus*. Available online at: http://www.saferoutesinfo.org/guide/index.cfm.

Pei, X., Wong, S.C., Sze, N.N. (2010). A joint-probability approach to crash prediction models. *Accident Analysis and Prevention* 43 (3), 1160-1166.

Peirson, J., Skinner, I., Vickerman, R.W. (1998). The microeconomic analysis of the external costs of road accidents. *Economica* 65, 429-440.

Pikora, T., Giles-Corti, B., Bull, F., Jamrozik, K., Donovan, R. (2003). Developing a framework for assessment of the environmental determinants of walking and cycling. *Social Science & Medicine* 56, 1693-1703.

Plaut, P.O. (2005). Non-motorized commuting in the US. *Transportation Research Part D* 10, 347-356.

Plummer, M., Best, N.G., Cowles, M.K., Vines, S.K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News* 6 (1), 7-11.

Pooley, C.G., Turnbull, J. (2000). Modal choice and modal change: the journey to work in Britain since 1890. *Journal of Transport Geography* 8, 11-24.

Porta, M. (2008). *A dictionary of epidemiology* (5th Edition). Oxford: Oxford University Press.

Prasad, K., 2007. What are relative risk, number needed to treat and odds ratio? *Annals of Indian Academy of Neurology* 10 (4), 225-230.

Pro Velo (2011). *Observatoire du vélo en Région de Bruxelles-Capitale – Rapport final 2010*. Pro Velo, Bruxelles. Available online at: [http://www.provelo.org/spip.php?article454]

Pucher, J., Buehler, R. (2006). Why Canadians cycle more than Americans: a comparative analysis of bicycling trends and policies. *Transport Policy* 13, 265-279.

Pucher, J., Buehler, R. (2008). Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. *Transport Reviews* 28 (4), 495-528.

Pucher, J., Buehler, R., Seinen, M. (2011). Bicycling renaissance in North-America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A* (in press).

Pucher, J., Dijkstra, L. (2003). Promoting safe walking and cycling to improve public health: lessons from the Netherlands and Germany. *Public Health Matters* 93 (9), 1509-1516.

Pucher, J., Dill, J., Handy, S. (2010). Infrastructure, programs, and policies to increase bicycling: An international review. *Preventive Medicine* 50, 106-125.

Pucher, J., Komanoff, C., Schimek, P. (1999). Bicycling renaissance in North America? Recent trends and alternatives policies to promote bicycling. *Transportation Research Part A* 33, 625-654.

Pulugurtha, S.S., Krishnakumar, V.K., Nambisan, S.S. (2007). New methods to identify and rank high pedestrian crash zones: An illustration. *Accident Analysis and Prevention* 39 (4), 800-811.

Quddus, M.A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accident Analysis and Prevention* 40, 1486-1497.

Rabl, A., de Nazelle, A. (2012). Benefits of shift from car to active transport. *Transport policy* 19, 121-131.

Raftery, A., Lewis, S. (1992). How many iterations in the Gibbs sampler? In Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.). *Bayesian Statistics*, Vol. 4, pages 763-774, Claredon Press, Oxford, UK.

Ramajo, J., Márquez, M.A., Hewings, G.J.D., Salinas, M.M. (2008). Spatial heterogeneity and interregional spillovers in the European Union: do cohesion policies encourage convergence across regions? *European Economic Review* 52, 551-567.

Räsänen, M., Summala, H. (1998). Attention and expectation problems in bicycle-car collisions: an in-depth study. *Accident Analysis and Prevention* 30 (5), 657-666.

Reynolds, C. CO, Harris, M.A., Teschke, K., Cripton, P.A., Winters, M. (2009). The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. *Environmental Health* 8 (47).

Rice, N. (2001). Binomial Regression. In Leyland, A.H., Goldstein, H. (Eds.), *Multilevel Modelling of Health Statistics*, pages 27-44, Wiley.

Rice, N., Jones, A. (1997). Multilevel models and health economics. *Health Economics* 6 (6), 561-575.

Richardson, A.J. (2000). *Seasonal and Weather Impacts on Urban Cycling Trips*. TUTI Report 1-2000. The Urban Transport Institute, Victoria, Australia.

Rietveld, P. (2000). The accessibility of railway stations: the role of the bicycle in The Netherlands. *Transportation Research Part D* 5, 71-75.

Rietveld, P. (2001). *Biking and walking: the position of non-motorized transport modes in transport systems*. Tinbergen Institute Discussion Paper TI 2001-111/3, the Netherlands.

Rietveld, P., Daniel, V. (2004). Determinants of bicycle use: do municipal policies matter? *Transportation Research Part A* 38, 531-550.

Rifaat, S.M., Tay, R., de Barros, A. (2011). Effect of street pattern on the severity of crashes involving vulnerable road users. *Accident Analysis and Prevention* 43, 276-283.

Ripley, B. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability* 13, 255-266.

Ripley, B. (1981). *Spatial statistics*. Wiley, Chichester, UK.

Rodgers, G.B. (1997). Factors associated with the crash risk of adult bicyclists. *Journal of Safety Research* 28 (4), 233-241.

Rodríguez, D.A., Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D* 9, 151-173.

Rojas-Rueda, D., de Nazelle, A., Tainio, M., Nieuwenhuijsen, M.J. (2011). Bike sharing system (Bicing) in Barcelona, Spain: a description and health impacts assessment. *British Medical Journal* 343, d425. doi:10.1136/bmj.d4521.

Saelens, B.E., Sallis, J.F., Frank, L.D. (2003). Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine* 25 (2), 80-91.

Schabenberger, O., Gotway, C.A. (2005). *Statistical methods for spatial data analysis*. Chapman & Hall/CRC, Florida, USA.

Schepers, J.P., Kroeze, P.A., Sweers, W., Wüst, J.C. (2011). Road factors and bicycle-motor vehicle crashes at unsignalized priority intersections. Accident Analysis and Prevention 43 (3), 853-861.

Schwenkglenks, M.M. (2007). *Multilevel modelling in the analysis of observational datasets in the health care setting*. Ph.D. Thesis in Sciences, University of Basel, Basel, Switzerland.

Shefer, D., Rietveld, P. (1997). Congestion and safety on highways: towards an analytical model. *Urban Studies* 34 (4), 679-692.

Shiode, S. (2008). Analysis of a distribution of point events using the network-based quadrat method. *Geographical Analysis* 40, 380-400.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall, London, UK.

Simon, D.S., 2001. Understanding the odds ratio and the relative risk. *Journal of Andrology* 22 (4), 533-536.

Smith, B.J. (2005). *Bayesian Output Analysis Program (BOA) Version 1.1 User's Manual*. University of Iowa College of Public Health, Department of Biostatistics, USA.

Spiegelhalter, D.J., Best, N., Carlin, B., van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B* 64, 583-639.

Spiegelhalter, D.J., Thomas, A., Best, N., Lunn, D. (2003). *WinBUGS user manual version 1.4*. Medical Research Council Biostatistics Unit, Cambridge, UK.

Spooner, P.G., Lunt, I.D., Okabe, A., Shiode, S. (2004). Spatial analysis of roadside Acacia populations on a road network using the network K-function. *Landscape Ecology* 19, 491-499.

SPW (Service Public de Wallonie) (2008). *Le Plan Escargot – un Soutien Financier aux Communes pour Favoriser les Modes Alternatifs de Déplacement*. La CeMathèque, Namur, Belgium.

Steenberghen, T., Dufays, T., Thomas, I., Flahaut, B. (2004). Intra-urban location and clustering of road accidents using GIS: a Belgian example. *International Journal of Geographical Information Systems* 18 (2), 169-181.

Steenberghen, T., Aerts, K., Thomas, I. (2010). Spatial clustering of events on a network. *Journal of Transport Geography* 18 (3), 411-418.

Stinson, M.A., Bhat, C.R. (2003). *An analysis of commuter bicyclist route choice using stated preference survey*. Transportation Research Board, Washington D.C, USA.

Stinson, M.A., Bhat, C.R. (2005). *A comparison of the route preferences of experienced and inexperienced bicycle commuters*. Transportation Research Board, Washington D.C., USA.

Sturtz, S., Ligges, U., Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software* 12 (3).

Thai, A., McKendry, I., Brauer, M. (2008). Particulate matter exposure along designated bicycle routes in Vancouver, British Columbia. *Science of the Total Environment* 405, 26-35.

Thill, J-C., Kim, M. (2005). Trip making, induced travel demand, and accessibility. *Journal of Geographical Systems* 7, 229-248.

Thisse, J-F., Thomas, I. (2010). Bruxelles au sein de l'économie belge : un bilan. *Regards Economiques* 80.

Thomas, A, Best, N., Lunn, D., Arnold, R., Spiegelhalter, D.J. (2004). *GeoBUGS user manual version 1.2*. Medical Research Council Biostatistics Unit, Cambridge University, UK.

Thomas, I. (1996). Spatial data aggregation: exploratory analysis of road accidents. *Accident Analysis and Prevention* 28, 251-264.

Thomas, I., Frankhauser, P., Vandenbulcke, G., Dujardin, C. (2011). Road accident while cycling: do urban morphologies really matter? Fractal evidences from Antwerp (Belgium). In: *Proceedings of the NECTAR Conference*, Antwerp, Belgium.

Tilahun, N.Y., Levinson, D.M., Krizek, K.J. (2007). Trails, lanes, or traffic: valuing bicycle facilities with an adaptative stated preference survey. *Transportation Research Part A* 41, 287-301.

Titheridge, H., Hall, P. (2006). Changing travel to work patterns in South East England. *Journal of Transport Geography* 14, 60-75.

Tobler, W.R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, 234-240.

Toint, P., Cornélis, E., Cirillo, C., Barette, P., Dessy, A., Jacobs, T., Verfaillie, T., Museux, J-M., Waeytens, E., Saelens, S., Durand, C., André, V., Van Hoof, K., Heylen, E., Pollet, I. (2001). *Enquête Nationale sur la Mobilité des Ménages : Réalisations et Résultats. Rapport Final*. Research contract MD/13/036 (final report) for Belgian Science Policy.

Tufféry, S. (2005). *Data mining and decisional statistics: the intelligence in the data bases*. Technip Editions, Paris, France.

Tzelgov, J., Henik, A. (1981). On the differences between Conger's and Velicer's definitions of suppressors. *Educational Measurement* 41, 1027-1031.

Tzelgov, J., Henik, A. (1991). Suppression situations in psychological research: definitions, implications and applications. *Psychological Bulletin* 109, 524-536.

Unger, R., Eder, C., Mayr, J.M., Wernig, J. (2002). Child pedestrian injuries at tram and bus stops. *Injury* 33 (6), 485-488.

Vandenbulcke, G., Steenberghen, T., Thomas, I. (2007). *Accessibility indicators to places and transports: Final Report*. Belgian Science Policy and FPS Mobility and Transports, Brussels, Belgium. Research project AP/02. Available at: [http://www.mobilit.fgov.be/fr/index.htm]

VanDerWal, J., Shoo, L.P., Graham, C., Williams, S.E. (2009). Selection pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling* 220, 589-594.

Van Hecke, E. (1998). Actualisering van de stedelijke hiërarchie in België. *Tijdschrift van het Gemeentekrediet van België*, No. 205, 45–76.

Van Malderen, L., Jourquin, B., Thomas, I., Vanoutrive, T., Verhetsel, A., Witlox, F. (2009). Mobility policies of the companies located in Belgium: are there success stories? In: *Proceedings of the BIVEC-GIBET Transport Research Day (3rd edition)*, Brussels, Belgium.

van Wee, B., Nijland, H. (2007). Health benefits of cycling. In: *CVS 2007 34th Edition*, Antwerpen, Belgium.

Vanoutrive, T., Van Malderen, L., Jourquin, B., Thomas, I., Verhetsel, A., Witlox, F. (2009). Let the business cycle! A spatial multilevel analysis of cycling to work. *Belgian Journal of Geography* 2, 217-232.

Vanoutrive, T., Van Malderen, L., Jourquin, B., Thomas, I., Verhetsel, A., Witlox, F. (2010). *From mobility management and multilevel modeling towards modeling mobility and multilevel management*. Ph.D. Thesis in Economic Sciences, University of Antwerp, Antwerp, Belgium.

Verhetsel, A. (1998). The impact of spatial versus economic measures in an urban transportation plan. *Computers, Environment and Urban Systems* 22 (6), 541-555.

Verhetsel, A., Thomas, I., Van Hecke, E., Beelen, M. (2007). *Monografie van het woonwerkverkeer in Belgïe*. Belgian Science Policy and FPS Mobility and Transports, Brussels, Belgium.

Verhetsel, A., Vanelslander, T. (2010). What location policy can bring to sustainable commuting: an empirical study in Brussels and Flanders, Belgium. *Journal of Transport Geography* 18, 691-701.

Walker, I. (2007). Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender. *Accident Analysis and Prevention* 39, 417-425.

Wang, Y., Nihan, N.L. (2004). Estimating the risk of collisions between bicycles and motor vehicles at signalized intersections. *Accident Analysis and Prevention* 36, 313-321.

Wang, C., Quddus, M.A., Ison, S.G. (2009). Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England. *Accident Analysis and Prevention* 41, 798-808.

Ward, J.H. (1963). Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association* 58 (301), 236-244.

Wardlaw, M.J. (2000). Three lessons for a better cycling future. *British Medical Journal* 321, 1582-1585.

Wardman, M., Tight, M., Page, M. (2007). Factors influencing the propensity to cycle to work. *Transportation Research Part A* 41, 339-350.

Warton, D.I., Shepherd, L. (2010). Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *Annals of Applied Statistics* 4 (3), 1383-1402.

Wedagama, D.M.P., Bird, R.N., Metcalfe, A.V. (2006). The influence of urban land-use on non-motorised transport casualties. *Accident Analysis and Prevention* 38, 1049-1057.

Wendel-Vos, G.C.W., Schuit, A.J., De Niet, R., Boshuizen, H.C., Saris, W.H.M., Kromhout, D. (2004). Factors of the physical environment associated with walking and bicycling. *Medicine and Science in Sports and Exercise* 36 (4), 725-730.

WHO (World Health Organization) (2002a). *30 minutes for a healthy life span.* Press release EURO/07/02, Copenhagen and Rome.

WHO (World Health Organization) (2002b). *A Physically Active Life through Everyday Transport, with a Special Focus on Children and Older People, and Examples and Approaches from Europe.* Regional Office for Europe, Copenhagen.

Winters, M., Davidson, G., Kao, D., Teschke, K. (2011). Motivators and deterrents of bicycling: comparing influences on decisions to ride. *Transportation* 38, 153-168.

Wintle, B.A. (2003). *Dealing with uncertainty in wildlife habitat models*. PhD. thesis, University of Melbourne (School of Botany), Australia.

Wintle, B.A., Bardos, D.C. (2006). Modeling species-habitat relationships with spatially autocorrelated observation data. *Ecological Applications* 16 (5), 1945-1958.

Wisz, M.S., Guisan, A. (2009). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecology* 9 (8).

Witlox, F., Tindemans, H. (2004). Evaluating bicycle-car transport mode competitiveness in an urban environment: an activity-based approach. *World Transport Policy and Practice* 10 (4), 32-42.

Woodcock, J., Banister, D., Edwards, P., Prentice, A.M., Roberts, I. (2007). Energy and transport. *The Lancet* 370, 1078-1088.

Woodcock, J., Edwards, P., Tonne, C., Armstrong, B.G., Ashiru, O., Banister, D., Beevers, S., Chalabi, Z., Chowdhury, Z., Cohen, A., Franco, O.H., Haines, A., Hickman, R., Lindsay, G., Mittal, I., Mohan, D., Tiwari, G., Woodward, A., Roberts, I. (2009). Public health benefits of strategies to reduce greenhouse-gas emissions: urban land transport. *The Lancet* 374 (9705), 1930-1943.

Xie, Z., Yan, J. (2008). Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems* 32, 396-406.

Yamada, I., Thill, J-C. (2004). Comparison of planar and network *K*-functions in traffic accident analysis. *Journal of Transport Geography* 12, 149-158.

Yamamoto, T., Hashijib, J., Shankar, V.N. (2008). Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. *Accident Analysis and Prevention* 40 (4), 1320-1329.

Yan, X., Ma, M., Huang, H., Abdel-Aty, M., Wu, C. (2011). Motor vehicle-bicycle crashes in Beijing : irregular maneuvers, crash patterns, and injury severity. *Accident Analysis and Prevention* 43 (5), 1751-1758.

Ye, F., Lord, D. (2011). *Investigating the effects of underreporting of crash data on three commonly used traffic crash severity models: multinomial logit, ordered probit and mixed logit models*. Transportation Research Record. Paper presented at the 90th Annual Meeting of the TRB.

Zahran, S., Brody, S.D., Maghelal, P., Prelog, A., Lacy, M. (2008). Cycling and walking: explaining the spatial distribution of healthy modes of transportation in the United States. *Transportation Research Part D* 13, 462-470.

Zaniewski, A.E., Lehmann, A., Overton, J.McC. (2002). Predicting species spatial distributions using presence-only data: a case-study of native New Zealand ferns. *Ecological Modelling* 157, 261-280.

Zarnetske, P.L., Edwards, T.C., Moisen, G.G. (2007). Habitat classification modeling with incomplete data: pushing the habitat envelope. *Ecological Applications* 17 (6), 1714-1726.

Zhu, L., Gorman, D.M., Horel, S. (2006). Hierarchical Bayesian spatial models for alcohol availability, drug "hot spots" and violent crime. *International Journal of Health Geographics* 5 (54).

Data sources and on-line resources

BRIC (Brussels Regional Informatics Center) – Brussels Urban Information System (UrbIS) database: http://www.cirb.irisnet.be/

BRIC – GeoLoc (orthophotos): http://geoloc.irisnet.be/

Brussels Mobility – IRIS II (Mobility Plan): http://www.bruxellesmobilite.irisnet.be/

Town of Brussels (mobility and public works) – Map of the "comfort area": http://www.bruxelles.be/artdet.cfm/4009

DGSEI (Directorate-General Statistics and Economic Information) – Road accidents (2006-2008) and 2001 socio-economic census: http://statbel.fgov.be/

EROS (Earth Resources Observation and Science) Center, Shuttle Radar Topography Mission (SRTM) – Elevation data set (Belgium): http://eros.usgs.gov/

Federal Planning Bureau – Transport database [in French]: http://www.plan.be/databases/database_det.php?lang=fr&TM=28&IS=60&DB=TRANSP&ID=14

Federal Public Service (FPS) Mobility and Transports – http://www.mobilit.fgov.be/

Google Earth – Aerial photography and satellite imagery (June 8th 2004, April 30th 2007, August 31st 2009): http://www.google.co.uk/intl/en_uk/earth/index.html

IBGE–BIM (Institut Bruxellois pour la Gestion de l'Environnement) – Data on motorized traffic (2006) for the Brussels Capital-Region: http://www.ibgebim.be/

SANET v.4 (Spatial Analysis on Networks), extension for ArcGIS 9.3 – Software for network analyses: http://sanet.csis.u-tokyo.ac.jp/

WinBUGS (Windows-based version of the BUGS software) – Software for Bayesian analyses: http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml

List of figures

Figure 1.1: Fatality rates for cyclists and bicycle share (%) in 2000 – A European comparison. EU 15 = European Union and its 15 former member countries 8
Figure 1.2: General objectives of the thesis (*i* & *ii*) within the contextual framework 10
Figure 1.3: General outline of the thesis 29
Figure 2.1: Percentage of commuters that use the bicycle as the only mode of transport 40
Figure 2.2: Risk of severe/fatal accident, defined as the average number of severe and fatal accidents (for commuter cyclists) per 100,000 bicycle-minutes, by municipality 43
Figure 2.3: Age and gender of cyclists versus commuting distance 47
Figure 2.4: Proportion of commuters that cycle in function of the urban hierarchy (H_j) of the workplace and commuting distances (2001) 48
Figure 2.5: Modal share for towns H_1 as destination, commuting (2001) 50
Figure 2.6: Modal share for towns H_2 as destination, commuting (2001) 50
Figure 2.7: Odds ratios for cycling (H_2 vs H_j) as a function of the travelled commuting distance 52
Figure 2.8: Classification of municipalities based on the two variables: bicycle use and risk of severe/fatal accident 56
Figure 2.9: Modal share of cyclists *versus* risk of severe/fatal accident: regional differences. Brussels is not illustrated here because of the small number of municipalities (19) 59
Figure 3.1: Commuters' modal share for distances up to 5 km (Belgium) 64
Figure 3.2: Percentage of commuter cyclists (women) having children being less than 5 years old, 6-11 years old, or 12-17 years old 72
Figure 3.3: OLS and ML residuals 87
Figure 3.4: Moran scatterplot and LISA cluster map for the spatial clustering of commuting by bicycle 90
Figure 3.5: Variation in bicycle use in Flanders as explanatory variables change. 94
Figure 3.6: Variation in bicycle use in Wallonia and Brussels as explanatory variables change 95

Figure 3.7: Absolute difference in commuter cycling between 1991 and 2001 .. 97
Figure 3.8: The residuals of the spatial regime specification 98
Figure 4.1: Evolution of the average number of cyclists (C), number of victims of cycling accidents (V), and ratio V/C 109
Figure 4.2: Centrographic measures for the distribution of (a) unreported cycling accidents (SHAPES survey) and (b) reported cycling accidents (DGSEI data) 140
Figure 4.3: Univariate spatial pattern analysis of both unreported (SHAPES) and reported (DGSEI) cycling accidents – Network *K*-function, Brussels' Pentagon and Second Crown 142
Figure 4.4: Locational tendency of unreported cycling accidents (SHAPES) with respect to reported cycling accidents (DGSEI) – Bivariate spatial pattern analysis, using the network cross *K*-function, Brussels' Pentagon and Second Crown 144
Figure 4.5: Locational tendency of both unreported (SHAPES) and reported (DGSEI) cycling accidents with respect to: (1) tram stops, and (2) on-road tram tracks – Bivariate spatial pattern analysis, using the network cross *K*-function and carried out at the scale of the First Crown 145
Figure 5.1: Modelling strategy 167
Figure 5.2: Data collection – conceptual framework 172
Figure 5.3: Spatial distribution of: (a) the exposure variable, i.e. the Potential Bicycle Traffic Index (PBTI), (b) control points, generated from the PBTI and drawn along the 'bikeable' network (without black zones) 175
Figure 5.4: Black spots of bicycle accidents (2006-2008) in the Brussels' European district. 176
Figure 5.5: Map of the predicted risk of having a cycling accident, computed from parameter estimates of the autologistic model 197
Figure 5.6: Predicted risk of having a cycling accident, separately computed for 4 risk factors: (a) tram tracks (on-road and crossings); (b) contraflow cycling (intersections are excluded); (c) van and truck traffic from 6 a.m. to 10:59 p.m. (all classes from 2 to 5); (d) autocovariate component. 198
Figure 6.1: Significant factors (and their interactions) influencing the risk of cycling accident in Brussels. 222

Appendix A.1: The urban hierarchy of Belgian municipalities 247
Appendix C.2: Blackspots of cycling accidents in the Pentagon 260

List of tables

Table 2.1: The means of variables in municipalities with different ranks in the urban hierarchy (H_j) 44
Table 2.2: Spearman and Pearson correlation coefficients between some selected variables (expected to be explanatories) and bicycle use as well as urban hierarchy (n = 589) 53
Table 3.1: Basic statistics and bivariate correlations with the proportion of commuting by bicycle at the scale of the municipalities (N = 589) 82
Table 3.2: Regression diagnostics for the OLS and ML estimations 84
Table 3.3: Regression coefficients for the OLS and ML estimations 85
Table 3.4: Regression diagnostics for the OLS and ML estimations, including the spatial regimes 91
Table 3.5: Regression coefficients for the spatial regime specification (ML estimation) 93
Table 4.1: Infrastructure factors (discrete) – Descriptive and comparative statistics 136
Table 4.2: Infrastructure factors (continuous) – Descriptive and comparative statistics 138
Table 4.3: Analysis of the spatial distribution of both unreported (SHAPES) and reported (DGSEI) cycling accidents, at the scale of 3 different subareas... 141
Table 4.4: Analysis of the spatial distribution of reported and unreported cycling accidents, with respect to infrastructure factors and using 3 spatial subareas (Pentagon, 1st and 2nd Crowns) 146
Table 5.1: Logistic (non-spatial) and auto-logistic models (spatial) – Results from the Bayesian framework 189
Table 5.2: Model-building specifications and model comparison 191
Table 6.1: Policy recommendations 230

Appendix B.1: Variables used: description, units of measurement and data sources 249
Appendix B.2: Regression coefficients for the spatial regime specification 252
Appendix B.3: Direct, indirect and total impact estimates, based on the spatial lag regime specification 253
Appendix C.1: Infrastructure factors – Description and data sources 257

Appendix D.1: List of risk factors 261
Appendix D.2: Descriptive statistics of the selected (discrete) risk factors.... 267
Appendix D.3: Descriptive statistics for the continuous risk factors 271
Appendix D.4: Logistic model – Results from the frequentist framework...... 274
Appendix D.5: Logistic model – Model fit and evaluation, diagnostics and inferential tests 275
Appendix D.6: Convergence diagnostics for the autologistic model........ 276
Appendix E.1: Approximate % of the time budget devoted to each task and chapter 280