New Methods for Modeling and Integrating Bicycle Activity and Injury Risk in an Urban Road Network

Jillian Strauss

Department of Civil Engineering and Applied Mechanics
McGill University, Montreal

December 2015

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Engineering.

© Jillian Strauss 2015
# Table of Contents

Table of Contents ......................................................................................................................... i  
List of Figures ................................................................................................................................. vii  
List of Tables ................................................................................................................................. ix  
Abstract ......................................................................................................................................... x  
Résumé ............................................................................................................................................. xii  
Acknowledgments ......................................................................................................................... xiv  
Contribution of Authors ............................................................................................................... xv  

## Chapter 1: Introduction ........................................................................................................... 1  
1.1. Context ................................................................................................................................. 2  
1.2. Problem Statement ............................................................................................................... 4  
1.3. Objectives ............................................................................................................................. 5  
  1.3.1. Specific Objectives ........................................................................................................... 5  
1.4. General Literature Review ............................................................................................... 6  
  1.4.1. Exposure Measures ........................................................................................................... 7  
  1.4.2. Geometric Design and Built Environment Characteristics ........................................... 8  
  1.4.3. Integral Bicycle Activity and Safety Modeling Approach .............................................. 10  
  1.4.4. Surrogate Safety Measures ............................................................................................... 11  
  1.4.5. Cyclist GPS Data ............................................................................................................... 13  
1.5. Original Contributions ....................................................................................................... 14  
1.6. General Methodology ......................................................................................................... 15  
1.7. Organization of the Document .......................................................................................... 19
Chapter 2: Cyclist Activity and Injury Risk Analysis at Signalized Intersections: A Bayesian Modeling Approach

2.1. Abstract

2.2. Introduction

2.3. Literature Review

2.4. Methodology

2.5. Data

2.6. Results

2.6.1. Factors Associated with Cyclist Injury Risk

2.6.2. Factors Associated with Bicycle Activity

2.6.3. Cyclist Injury Risk along Corridors

2.7. Conclusion and Future Work

2.8. Addendum

2.9. Acknowledgments

2.10. References

Link between Chapter 2 and Chapter 3

Chapter 3: Multimodal Injury Risk Analysis between Road Users at Signalized and Non-Signalized Intersections

3.1. Abstract

3.2. Introduction

3.3. Literature Review

3.4. Methodology

3.5. Site Selection and data

3.5.1. Site Selection

3.5.2. Signalized and Non-Signalized Intersection Inventories
Link between Chapter 3 and Chapter 4 ................................................................. 72

Chapter 4: Mapping Cyclist Activity and Injury Risk in a Network Combining Smartphone GPS Data and Bicycle Counts ................................................................. 73

4.1. Abstract ..................................................................................................................... 74
4.2. Introduction ............................................................................................................... 75
4.3. Literature Review ...................................................................................................... 77
4.4. Methodology ............................................................................................................. 79
  4.4.1. Assign GPS Traces to the Network ................................................................. 80
  4.4.2. AADB Volumes from Counts and Extrapolation Functions .................... 80
  4.4.3. Validation of Predicted AADB ............................................................... 82
  4.4.4. Safety Applications ...................................................................................... 82
4.5. Data ......................................................................................................................... 83
  4.5.1. Smartphone GPS Trips and Traces ............................................................ 83
  4.5.2. Short-Term and Long-Term Counts ............................................................ 83
  4.5.3. Cyclist Injuries ................................................................................................. 85
4.6. Results ..................................................................................................................... 85
  4.6.1. AADB Estimation ...................................................................................... 86
  4.6.2. Risk Analysis ................................................................................................. 89
4.7. Conclusion and Future Work ................................................................. 93
4.8. Acknowledgments ............................................................................. 95
4.9. Addendum .......................................................................................... 96
4.10. References ....................................................................................... 96

Link between Chapter 4 and Chapter 5 ..................................................... 99

Chapter 5: Cyclist Deceleration Rate as Surrogate Safety Measure in Montreal using Smartphone GPS Data ............................................................................. 100
5.1. Abstract ............................................................................................. 101
5.2. Introduction ......................................................................................... 102
5.3. Literature Review ................................................................................ 103
5.4. Methodology ...................................................................................... 105
  5.4.1. Cleaning Rules ............................................................................ 106
  5.4.2. Separate Intersections and Segments ....................................... 107
  5.4.3. Compute Instantaneous Speeds ................................................ 108
  5.4.4. Compute Instantaneous Accelerations and Decelerations .......... 109
  5.4.5. Empirical Bayes (EB) Estimator ............................................. 110
  5.4.6. Explore the Relationship between Injuries and Deceleration Rate 111
5.5. Site Selection and Data ....................................................................... 111
  5.5.1. Smartphone GPS Trips and Traces ......................................... 111
  5.5.2. Historical Cyclist Injury Data .................................................. 112
5.6. Results ................................................................................................ 113
  5.6.1. Intersections .............................................................................. 113
  5.6.2. Road Segments ......................................................................... 115
  5.6.3. Corridors ................................................................................... 118
5.7. Conclusion and Future Work ............................................................... 118
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.8.</td>
<td>Acknowledgements</td>
<td>119</td>
</tr>
<tr>
<td>5.9.</td>
<td>References</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td><strong>Link between Chapter 5 and Chapter 6</strong></td>
<td><strong>122</strong></td>
</tr>
<tr>
<td><strong>Chapter 6:</strong></td>
<td>Speed, Travel Time and Delay for Intersections and Road Segments in Montreal using Cyclist Smartphone GPS Data</td>
<td><strong>123</strong></td>
</tr>
<tr>
<td>6.1.</td>
<td>Abstract</td>
<td>124</td>
</tr>
<tr>
<td>6.2.</td>
<td>Introduction</td>
<td>125</td>
</tr>
<tr>
<td>6.3.</td>
<td>Literature Review</td>
<td>126</td>
</tr>
<tr>
<td>6.4.</td>
<td>Methodology</td>
<td>128</td>
</tr>
<tr>
<td>6.4.1.</td>
<td>Distance, Time and Speed Computations</td>
<td>130</td>
</tr>
<tr>
<td>6.4.2.</td>
<td>Data Filtering</td>
<td>130</td>
</tr>
<tr>
<td>6.4.3.</td>
<td>Speed Smoothing</td>
<td>132</td>
</tr>
<tr>
<td>6.4.4.</td>
<td>Separate Segments and Intersections</td>
<td>133</td>
</tr>
<tr>
<td>6.4.5.</td>
<td>Average Speed on Segments</td>
<td>134</td>
</tr>
<tr>
<td>6.4.6.</td>
<td>Travel Time through Intersections</td>
<td>134</td>
</tr>
<tr>
<td>6.4.7.</td>
<td>Intersection Delay</td>
<td>135</td>
</tr>
<tr>
<td>6.4.8.</td>
<td>Segment Speed Model</td>
<td>136</td>
</tr>
<tr>
<td>6.5.</td>
<td>Site Selection and Data</td>
<td>136</td>
</tr>
<tr>
<td>6.5.1.</td>
<td>Smartphone GPS Trips and Traces</td>
<td>136</td>
</tr>
<tr>
<td>6.5.2.</td>
<td>Weather Data</td>
<td>137</td>
</tr>
<tr>
<td>6.6.</td>
<td>Results</td>
<td>137</td>
</tr>
<tr>
<td>6.6.1.</td>
<td>Segment Speed</td>
<td>137</td>
</tr>
<tr>
<td>6.6.2.</td>
<td>Intersection Delay</td>
<td>139</td>
</tr>
<tr>
<td>6.6.3.</td>
<td>Segment Speed Model</td>
<td>141</td>
</tr>
<tr>
<td>6.6.4.</td>
<td>Applications</td>
<td>144</td>
</tr>
</tbody>
</table>
6.7. Conclusion and Future Work ................................................................. 147
6.8. Acknowledgments ............................................................................. 148
6.9. References ......................................................................................... 149

Chapter 7: Conclusion and Future Work ...................................................... 151
7.1. General Conclusion ........................................................................... 152
7.2. Future Work ....................................................................................... 155
7.3. Publications ......................................................................................... 156

Appendix A: Non-Signalized Intersections ............................................... 159
A.1. Results ............................................................................................... 159
   A.1.1. Factors Associated with Cyclist Injury Risk .................................. 159
   A.1.2. Factors Associated with Bicycle Activity ...................................... 159

Appendix B: Additional Multimodal Results ............................................. 161
B.1. Expected Number of Injuries and Risk ............................................... 161
   B.2 Correlations between Flows, Injuries and Risk .................................. 164

Appendix C: Map-Matching, Additional Safety Application and GPS Data .... 166
   C.1 Map-Matching ................................................................................... 166
   C.2 Injury Maps ....................................................................................... 166
   C.3 Safety Application ............................................................................ 169
   C.4 Smartphone GPS Trips and Traces .................................................. 171
List of Figures

Figure 1-1 Conceptual Framework ................................................................. 19
Figure 2-1 Study intersections and average observed injury rates .................. 33
Figure 2-2 Injury rates along corridors ........................................................ 42
Figure 3-1 Sample of Signalized (green circles) and Non-Signalized (red squares) Intersections .................................................................................................................. 58
Figure 3-2 Distribution of Injuries at Signalized (left) and Non-Signalized (right) Intersections 60
Figure 3-3 Percentage of Total Observed and Expected (in brackets) Injuries by Mode at Signalized (left) and Non-Signalized Intersections (right) .......................................................... 65
Figure 3-4 Distribution of Intersections According to Risk ............................. 66
Figure 3-5 Box Plots for Different Levels of Motor-Vehicle Flow for Signalized (left) and Non-Signalized (right) Intersections .................................................. 67
Figure 4-1 Flowchart of Methodology ............................................................. 79
Figure 4-2 Segment and Intersection Buffers and GPS Points ........................ 80
Figure 4-3 Intersections with Manual Counts and Automatic Count Locations .......................................................... 84
Figure 4-4 Raw Injuries at Intersections (left) and along Segments (right) ........ 85
Figure 4-5 Predicted AADB at Intersections (left) and Segments (right) ....... 88
Figure 4-6 AADB and EB Risk at Signalized Intersections (top: a and b), Non-Signalized Intersections (middle: c and d) and Segments (bottom: e and f) .................. 92
Figure 4-7 Box Plots of EB Injuries and EB Risk at Intersections .................. 93
Figure 5-1 Methodology Steps ........................................................................ 106
Figure 5-2 Sample of the Road Network showing the Separation of Intersections and Segments .................................................................................................................. 108
Figure 5-3 Sample of Acceleration/Deceleration Data Showing the Dangerous Threshold ..... 110
Figure 5-4 Speed Distributions for Segments and Intersections ..................... 113
Figure 5-5 Heat Maps of Dangerous Decelerations (top left), EB Injuries (top right), Dangerous Decelerations per Trip (bottom left) and EB risk (bottom right) at Signalized and Non-Signalized Intersections .................................................................................................................. 115
Figure 5-6 Heat Maps of Dangerous Decelerations, EB Injuries and Rates on Segments ....... 116
Figure 5-7 Number of EB Injuries per Km (top) and Number of Dangerous Decelerations per Km (bottom) for Corridors
Figure 6-1 Methodology Flowchart
Figure 6-2 Sample of Speed Filtering Technique
Figure 6-3 Speed Distributions for Segments and Intersections
Figure 6-4 Average Speed using Median 3 Filter on All Road Segments Downtown (top) and Zoom-In (bottom)
Figure 6-5 Sample of Intersection Delays per Approach (top) and Averaged for all Intersections Downtown (bottom)
Figure 6-6 Shortest Routes Based on Different Levels of Preference for Cycle Tracks
Figure 6-7 Shortest Route, Fastest Route and Safest Route
Figure B-1 Number of Injuries and Risk at Signalized (a and b) and Non-Signalized (c and d) Intersections
Figure C-1 EB Injuries and Risk at Signalized Intersections (Top), Non-Signalized Intersections (Middle) and Segments (Bottom)
Figure C-2 Top 25 Most Dangerous Signalized Intersections, Non-Signalized Intersections and Segments Based on EB Injuries (Top) and EB Risk (Bottom)
Figure C-3 Daily and Cumulative Number of Cyclists Registered
Figure C-4 Daily and Cumulative Number of Trips
Figure C-5 Raw GPS Points (left) and Traces (right)
List of Tables

Table 2-1 Summary statistics ........................................................................................................ 34
Table 2-2 Results ............................................................................................................................ 37
Table 2-3 Ranking of corridors .................................................................................................... 43
Table 3-1 Summary Statistics .................................................................................................... 59
Table 3-2 Cyclist, Pedestrian and Motor-Vehicle Injury Model Results for Signalized and Non-
Signalized Intersections ........................................................................................................ 64
Table 3-3 Summary Statistics of Expected Risk ........................................................................ 65
Table 4-1 AADB Model Results .................................................................................................. 87
Table 4-2 Injury Model Results to Compare AADB from Manual Counts to AADB from GPS
Trips for Validation Purposes .................................................................................................... 88
Table 4-3 Injury Model Results for Signalized Intersections, Non-Signalized Intersections and
Segments ......................................................................................................................................... 91
Table 5-1 Summary of Accident Data ......................................................................................... 112
Table 5-2 Summary of Speed and Acceleration/Deceleration for Intersections and Segments . 113
Table 5-3 Summary Statistics for Signalized and Non-Signalized Intersections ....................... 114
Table 5-4 Summary Statistics for Road Segments ........................................................................ 116
Table 6-1 Average Unfiltered and Filtered Speeds by Road Type ............................................... 139
Table 6-2 List and Description of Variables Tested in the Speed Model ..................................... 141
Table 6-3 Comparison of Results using Unfiltered and Filtered Speeds .................................... 142
Table 6-4 Model Results with One Personality Variable .............................................................. 143
Table 6-5 Model Results with Personality and Other Variables ............................................... 144
Table A-1 Non-Signalized Intersection Results ........................................................................ 160
Table B-1 Correlation Matrices for Signalized and Non-Signalized Intersections ..................... 165
Table C-1 Correlation between Raw and Map-Matched GPS Counts for Different Facility Types
........................................................................................................................................................ 166
Table C-2 Summary Statistics of Trip and User Attributes .......................................................... 172
Abstract

Despite the fact that the bicycle modal split is low, an increase in urban cycling activity has started to appear in many urban areas in Canada and the United States. However, as bicycle flows increase so do concerns for cyclist safety which have been pushed to the forefront. In this context, the need for new planning tools and data collection methods to investigate bicycle activity and safety have emerged in the field of transportation engineering. While this field of research continues to grow, several important gaps remain in the literature. Studies focusing on investigating the link between cyclist safety, geometric design and built environment characteristics as well as traffic conditions are rare in the current literature. Even rarer are studies focused on investigating safety across different facilities, for the different road users (cyclists, pedestrians and motor-vehicles) who share these facilities and the factors affecting their injury occurrence. Many tools have been developed to obtain cyclist volumes combining short- and long-term counts at specific sites. However, tools to estimate bicycle volumes at the entire network level are missing. All previous safety studies have considered a sample of sites and have not been able to compute and map risk in the entire network of intersections and road segments. Recently, road safety research has been interested in developing surrogate safety measures to identify injury factors or dangerous locations. Dangerous decelerations as a surrogate safety measure for cyclists and their correlation with accidents has yet to be investigated for the entire network. To date little is known about cyclist speeds and travel times along segments and delays through intersections at the disaggregate level for the entire network.

In order to address these limitations, the general objective of this thesis is to propose new methods to model and estimate bicycle activity and injury risk at different spatial levels (site, corridor, and network levels) combining different sources of data. The proposed methods are then used to identify risk factors as well as to map risk indicators based on accidents and hard braking along corridors and at intersections in the entire network. More specifically, the objectives of this thesis are, to: 1) develop a Bayesian modeling framework to simultaneously model injury and activity outcomes for cyclists and study the role of geometric design and built environment characteristics on both outcomes, 2) carry out a comparative analysis between the injuries, levels of flow and risk for the three modes at signalized and non-signalized intersections and investigate the impact of vehicle traffic on the safety of non-motorized modes, 3) improve current methods
for determining bicycle exposure measures by combining manual counts, automatic counts and GPS trip data to estimate and map bicycle flows, injuries and risk throughout the entire network of intersections and road segments, 4) develop a methodology to obtain deceleration rate for cyclists at intersections and segments using GPS data, explore the relationship between observed injuries and deceleration rate (DR) and validate DR as a surrogate safety measure, and 5) develop a methodology to estimate cyclist speeds, travel times and delays in a road network using GPS data and identify the factors affecting cyclist speeds along segments.

Among the contributions, this thesis proposed a modeling framework to simultaneously study cyclist injury occurrence and bicycle activity to overcome the issues of endogeneity. Cyclist injury occurrence was found to increase with increasing bicycle and vehicle traffic flows, specifically vehicle turning volumes. The presence of bus stops at intersections and crossing lengths were also found to increase cyclist injury occurrence. On the other hand, the presence of a raised median was identified as a factor reducing cyclist injury occurrence. Using the developed models, cycling corridors were defined and ranked based on the posterior expected number of injuries and injury rates. This allows the identification of not only dangerous intersections but also of dangerous corridors. A similar methodology was then applied from a multimodal perspective, to study the activity and safety outcomes for cyclists, pedestrians and motor-vehicle occupants. Similar to the cyclist injury results, pedestrian injury occurrence was found to increase with increasing pedestrian and vehicle flows. This analysis also emphasized the risk imposed by vehicles on cyclists and pedestrians. This thesis expanded on the previous safety work by developing and applying a methodology to combine short and long-term count data with a new source of bicycle GPS data, from a Smartphone application, to compute and map average annual daily bicycle flows and cyclist risk of injury throughout the entire network of intersections and road segments. The results identify that bicycle flow is greatest in the central neighbourhoods of the island of Montreal whereas risk is greatest outside these central neighbourhoods, which is also where bicycle infrastructure is lacking. Using the smartphone GPS data, hard deceleration data was extracted and proposed as a surrogate safety measure and its correlation with accidents was evaluated. Finally, a methodology to compute speeds along segments and delays through intersections was proposed based on the GPS data. The results from this thesis were then combined into a routing concept capable of identifying the safety, shortest and fastest routes based on each cyclist’s preferences.
Résumé

Bien que la fraction des déplacements effectués à vélo soit encore faible, l’usage du vélo augmente désormais dans les villes des États-Unis et du Canada. Dans ce contexte, de nouvelles méthodes de collecte de données et de nouveaux outils de planification ciblant l’usage et la sécurité du vélo urbain ont émergé. Néanmoins, les connaissances actuelles comportent d’importantes lacunes. Peu de recherches ont investigué l’association entre la sécurité des cyclistes, la géométrie routière, l’environnement bâti et la circulation des véhicules. Les facteurs influençant la sécurité de différents usagers de la route (piétons, cyclistes, occupants de véhicules à moteur), dans différents contextes, restent peu connus. Plusieurs outils ont été développés pour estimer les volumes de cyclistes à des sites spécifiques, cependant aucun outil ne permet d’estimer les volumes de cyclistes sur l’ensemble d’un réseau routier. Les recherches antérieures sont basées sur des échantillons de convenance et n’ont pas pu estimer ou cartographier le risque de blessures sur l’ensemble d’un réseau routier, incluant intersections et segments routiers. La décélération des cyclistes sur le réseau routier pourrait être un indicateur indirect de leur sécurité (∈surrogate safety measure⊥, mais son association avec les collisions et les blessures n’a pas encore été validée. Enfin, la vitesse et la durée des déplacements à vélo sur les segments routiers ainsi que les délais aux intersections sont peu connus.

L’objectif général de cette thèse est de proposer de nouvelles méthodes pour modéliser et estimer l’activité cycliste et le risque de blessures à différentes échelles spatiales, en intégrant différentes sources de données. Les méthodes proposées permettront d’identifier les facteurs de risque environnementaux et de cartographier le risque de blessures, basé sur les accidents et les indicateurs indirects (ex. décélérations), à l’échelle d’intersections, de corridors et de l’ensemble du réseau routier. Les objectifs spécifiques de cette thèse sont : 1) développer un modèle d’analyse Bayésien pour étudier simultanément les déterminants de l’activité cycliste et du risque de blessures, notamment l’influence de l’environnement bâti et de la géométrie des routes; 2) réaliser une analyse comparative de l’exposition, du nombre et du risque de blessures pour trois modes de transport, aux intersections signalisées et non-signalisées, et explorer l’impact de la circulation automobile sur la sécurité des usagers non-motorisés; 3) améliorer les méthodes actuelles permettant d’estimer l’exposition des cyclistes, en combinant des comptages manuels, des comptages automatiques et des données GPS sur les déplacements, afin d’estimer et de
cartographier l’activité cycliste et le risque de blessures sur l’ensemble du réseau routier; 4) développer une méthode pour obtenir le taux de décélération des cyclistes sur les segments routiers et aux intersections, en utilisant des données GPS sur les déplacements, et explorer son association avec les blessures observées; 5) développer une méthode basée sur des données GPS pour estimer la vitesse des cyclistes, la durée des déplacements et les délais, et identifier les facteurs influençant la vitesse des cyclistes sur les segments routiers.

Cette thèse propose un cadre d’analyse pour étudier simultanément l’activité cycliste et le risque de blessures et ainsi surmonter l’enjeu de l’endogénéité. Parmi les résultats observés, l’incidence des cyclistes blessés augmente avec les volumes de cyclistes et de véhicules, notamment avec le volume de véhicules effectuant un virage. De plus, le nombre de cyclistes blessés augmente avec la présence d’arrêt d’autobus et la longueur de la traversée mais diminue avec la présence de terre-pleins. Des corridors cyclistes ont aussi été définis et ont été classés, triés, selon le nombre et le taux de cyclistes blessés. Une méthodologie similaire a été utilisée dans une perspective multimodale, afin d’étudier l’exposition et le risque de blessures pour les cyclistes, les piétons et les automobilistes. Selon ces analyses, une augmentation du volume de véhicules est associée à une augmentation du nombre de piétons et de cyclistes blessés mais aussi – en tenant compte des volumes de marche et de vélo - à une augmentation du risque de blessures pour les piétons et cyclistes.

Cette thèse développe et applique une méthodologie pour calculer et cartographier le débit journalier annuel moyen de cyclistes ainsi que le risque de blessures sur l’ensemble du réseau routier, en intégrant des comptages avec une nouvelle source de données GPS sur les déplacements cyclistes, provenant de téléphones intelligents (« smartphones »). Selon ces analyses, les volumes de cyclistes sont plus élevés dans les quartiers centraux tandis que le risque de blessure est plus élevé en dehors de ces quartiers centraux, des lieux où il y a peu d’infrastructures cyclistes. Les données GPS ont aussi été utilisées pour mesurer les décélérations subites des cyclistes, et cet indicateur indirect du risque d’accident a été corrélé avec le nombre estimé de cyclistes blessés. Enfin, une méthodologie basée sur les données GPS a été développée pour calculer la vitesse des cyclistes sur les segments routiers et les délais aux intersections. Les résultats de cette thèse ont été intégrés pour identifier les trajets cyclistes les plus sécuritaires, les plus courts et les plus rapides.
Acknowledgments

I want to thank my supervisors, Luis Miranda-Moreno and Patrick Morency, for their guidance and support. With their help and guidance we were able to develop such interesting and important research topics for this thesis.

I want to acknowledge and thank the following people and companies, each of whom played an important role in my research:

Sohail Zangenehpour for being available when I needed to share and discuss research ideas and methodologies.

All the McGill students, including my brother Saul, who helped to collect the intersection inventory data.

The Montreal Department of Public Health (Direction de la Santé Publique) for providing the injury data.

Brisk Synergies and the city of Montreal for giving me access to the Mon RésoVélo GPS data.

and

Fonds québécois de la recherche sur la nature et les technologies (FQRNT), for their financial support.

Last but not least, I would very much like to extend a big heartfelt thanks to my family, specifically my parents, David and Laurie, for always supporting me in my research endeavors.
Contribution of Authors

Please note that this is a manuscript-based thesis consisting of five journal papers. These papers were written in collaboration with other authors. The titles of the articles, names of the authors, and the names of the journals are listed below. It is worth mentioning that the author of this thesis is the sole student, among the co-authors, who was responsible for conducting the research, analyzing the data and preparing the manuscripts. The papers making up the first three chapters of this thesis have already been published, the fourth one has been submitted for publication and the last one will be submitted shortly. The author’s supervisors, Prof. Luis Miranda-Moreno and Patrick Morency, provided guidance and editorial revisions throughout the entire process. Sohail Zangenehpour and Nicolas Saunier also provided guidance on one paper. Minor edits have been made to the original published papers.


5. Strauss, J. and L.F. Miranda-Moreno. Speed, Travel Time and Delay for Intersections and Road Segments in Montreal using Cyclist Smartphone GPS Data. Accepted for presentation at the 95th Transportation Research Board Annual Meeting and to be submitted for publication in Transportation Research Part C: Emerging Technologies.
Chapter 1

Introduction
Chapter 1: Introduction

1.1. CONTEXT

Due to growing concerns for climate change as well as personal health, cities have begun to witness a shift away from motorized travel to non-motorized modes (Commission of the European Communities, 2005; National Center for Chronic Disease Prevention and Health Promotion, 2009; Pucher et al., 2011a; Rutter et al., 2008; Vélo Québec, 2011a; Ville de Montreal, 2010; World Health Organization, 2004). In Canada, from 1996 to 2006, the number of daily bicycle commuters increased by 42% (Pucher et al., 2011a). With the growing number of cyclists, comes an increasing concern for their safety. At the same time, cities are striving to achieve sustainable transportation, so the push towards non-motorized modes is even greater. More cyclists in urban environments is likely to increase the absolute number of injuries and fatalities unless changes are implemented to the physical environment, specifically to the geometric design and built environment (Elvik, 2009; Fernandes, 2013; Strauss, 2012a). Fortunately, local governments are also seeking projects and initiatives to provide safer and more comfortable road environments for non-motorized modes of transportation. Accordingly, many cities in North America have implemented or planned many actions and countermeasures including the redesign of road space in the transportation network (installation of bicycle facilities, pedestrian zones, upgrading intersections, etc.), modification of the built environment (densification and diversification of land use) and redesign of traffic controls (signal controls at signalized intersections, new pavement makings, etc.). As a prime example, the city of Montreal continues to expand and upgrade its bicycle facility network. In 2000, there were about 300 kilometres of bicycle infrastructure which had been expanded to 500 kilometres by 2010 (Vélo Québec, 2011b, 2001). Despite the many efforts and projects implemented in North American cities, road safety remains a major concern and acts as a barrier for the development of non-motorized transportation (Transport Canada, 2010). In Canada, traffic-related accidents are the fourth leading cause of death and the leading cause of death for Canadians under the age of 44 (Public Health Agency of Canada, 2012). On average, almost 2,200 people, car drivers and passengers, motorcyclists, cyclists and pedestrians are killed each year on Canadian roads. Of these 2,200, on average, 54 of these are cyclists who are killed every year, representing about 2.5% of
fatalities in Canada. In 2012, 43% of road fatalities and 76% of injuries occurred in Canadian urban areas, of which 3% of fatalities and 4% of serious injuries were cyclists (Transport Canada, 2014).

The modal share of cycling remains very low. The Canadian Census began collecting bicycle data in 1996 when the modal share of cycling to work was at 1.1%. The modal share increased and reached 1.2% in 2001 and then reached 1.3% in 2006 (Statistics Canada, 2006, 2001). While the modal share of cycling for work trips, 1.3% in 2006, is still low in comparison to other modes, 80% by car, 11% by public transit and 6.4% by walking, 1.3% by other, it is continuing to increase. Especially in Canada's big cities, Toronto, Vancouver and Montreal, the modal shares of cycling remain above the national average. The modal shares of cycling increased, from 1.0% to 1.2% in Toronto, from 1.7% to 1.8% in Vancouver and from 1.6% to 1.7% in Montreal, from 2006 to 2012 (Statistics Canada, 2011). As the modal share of cycling continues to rise without any considerations made to accommodate more cyclists on the road, more road injuries are expected and therefore bringing issues of cyclist safety to the forefront. Studies focusing on data collection techniques to obtain geometric design and built environment characteristics as well as exposure data and modeling approaches to identify the factors affecting cyclist safety in urban environments are needed to design and provide safe environments for cyclists to reach their destinations.

Unfortunately cyclists fall into the category of vulnerable road users. Should a collision occur between a cyclist and a motor-vehicle, where injuries are sustained, chances are much greater that the cyclist is seriously injured or even killed where the motor-vehicle occupants may remain unharmed. When a collision occurs between a motor-vehicle and a cyclist, 100 out of 1,000 times, this requires the hospitalization of the cyclist, and 10 out of 1,000 times this results in the death of the cyclist. For collisions involving cyclists but without motor-vehicle involvement, 50 out of 1,000 times this requires the hospitalization of the cyclist and less than 1 out of 1,000 times results in a fatality (Vélo Québec, 2006). On the island of Montreal from 1999 to 2008, almost 70,000 road users were injured of which 13% were cyclists. In Montreal over the five years from 2005 to 2009, 3 cyclists were killed, 36 sustained major injuries needing hospitalisation and 651 cyclists sustained minor injuries (Vélo Québec, 2011b). Furthermore, amongst the road elements, intersections are critical, accounting for 60% of the total injuries in cities such as Montreal. Urban intersections are complex elements in the urban road network where interactions occur between motor-vehicles, cyclists and pedestrians, with a small proportion leading to collisions. Considering the importance of safety, transportation agencies are constantly screening the road network to identify locations,
corridors, or any other area for which safety countermeasures can be implemented. Once sites are selected and countermeasures are implemented, determining their impact, in terms of injury reduction, is a required task to evaluate their effectiveness. Mapping injury risk and traffic intensity throughout the network over time is an important challenge, raising the need for data collection, monitoring, integration and modeling tools that can help in the road safety process (road screening, countermeasure identification and evaluating the impact of interventions). Road safety research is a continuously expanding field. Emerging studies focus on the effects of both motorized and non-motorized exposure measures, geometric design and built environment characteristics, weather conditions, bicycle infrastructure as well as driver behaviour and many other aspects which can have an impact on road safety.

1.2. PROBLEM STATEMENT

As cities strive to adapt themselves towards sustainable cities, with less dependence on motorized modes of transportation and enhanced motivation towards the use of active modes, such as walking and cycling, new tools and methods are needed to fill the gaps with respect to data collection, monitoring and integration methods as well as modeling approaches. There is a need for new methods for modeling and integrating bicycle activity and injury risk in the planning, design and operation of urban transportation systems. Currently, there is a lack of systematic methods for collecting and integrating traffic exposure measures for an entire network or population of intersections both signalized and non-signalized as well as road segments. To date, most studies in the literature focusing on cyclist safety have used small samples of sites, have focused solely on signalized intersections and have relied on manual cyclist and motor-vehicle counts for only a few hours and without correcting for temporal trends. As mentioned, to overcome these limitations, new methods for integrating bicycle activity and injury risk are needed at the network level. In order to attain sustainable cities, key elements that must be addressed are urban mobility and safety for all modes of transportation and their interactions. Also, little is known about the contributing factors associated with the risk across road user types (cyclist, pedestrian and motor-vehicle) and across facility types (signalized and non-signalized intersections as well as road segments).

This research also aims to consider cyclist safety and mobility in an urban environment which represents a key element needed for a multimodal road safety approach. This approach can help build comprehensive cyclist safety portraits where the risk at each facility can be determined.
This work will focus on obtaining reliable exposure data, developing safety performance functions, computing injury risk, extracting speeds and travel times, for cyclists for all facility types throughout the entire network. In other words, this work will provide in-depth analysis for cyclists in an urban area with a minor focus on the safety of pedestrians and motor-vehicle occupants as well.

This research also investigates the link between surrogate measures and traditional crash-based outcomes. We hypothesize that surrogate safety measures are capable of providing a complementary approach to safety with respect to accident data. Validating surrogate safety measures has great safety implications since injuries can be prevented and potentially lives can be saved. In other words, dangerous locations in the network can be identified and treated without the site ever witnessing any injuries. The traditional approach is to wait several years to observe a sufficient number of injuries and then identify the faults in road design and geometry and only then make changes, which may be too late for the cyclists.

1.3. OBJECTIVES

In response to the existing shortcomings in the literature, the general objective of this thesis is to propose new methods for identifying factors affecting cyclist safety across different facilities and map bicycle activity and injury risk throughout the entire network using traditional and surrogate safety measures. This work provides methods and empirical evidence that help to compare activity levels and risk among all sites which can serve great purpose for urban cities. This can help guide the decision of where new infrastructure is needed or where to expand current infrastructure. Also knowledge of where cyclists are riding and the risk of injury they are exposed to, can help the city best select sites in need of safety interventions and the appropriate countermeasures. To illustrate the methods proposed in this research, the island of Montreal, Quebec is used as the application environment.

1.3.1. Specific Objectives

The specific objectives of this thesis are, to:

1) Propose a Bayesian modeling framework to simultaneously model cyclist injury occurrence and bicycle activity levels at signalized intersections. This methodology is capable of identifying the effects that traffic flow and geometry and built environment characteristics
have on both outcomes. An extensive inventory was built for the implementation of the proposed methodology.

2) Extend the Bayesian modeling framework to a multimodal modeling approach to estimate the injury risk for all road users (cyclists, pedestrians and motor-vehicle occupants) and intersection types (signalized and non-signalized intersections). A comparative analysis across modes and two facility types (signalized and non-signalized intersections) is developed to investigate the levels of risk across modes and the contributing factors for each facility type.

3) Expand upon current methods based on point-based measures into network-level measures to map bicycle activity in the entire network by combining manual counts, automated counts and GPS trip data. This methodology proposes an extrapolation function to convert cyclist GPS traces to average annual daily bicycle flows (AADB) throughout the entire network of intersections and road segments. The AADB values can then be used to map injury risk throughout the entire network.

4) Propose a surrogate safety methodology based on cyclist deceleration rates using GPS data in order to quantify the number of dangerous decelerations (hard braking). This surrogate safety measure can then be mapped in the entire network and its relationship with observed injuries can be explored for validation purposes.

5) Develop a methodology to estimate cyclist speeds and delays through signalized intersections in the network and identify the factors affecting cyclist speed along segments and account for each cyclist's personality. Combining the results obtained from the previous objectives, routes from origin to destination can be provided based on different criteria including the developed safety indicators and travel speeds.

1.4. GENERAL LITERATURE REVIEW

The existing and emerging road safety literature is extensive and has dealt with many issues including data collection, extrapolation and modeling approaches. More specifically, the literature has sought to develop methods to obtain cyclist exposure measures from combining short- and long-term counts (Miranda-Moreno et al., 2013), and predicting flows from regression models (Griswold et al., 2011). A class of studies have modeled cyclist accident and injury occurrence but multimodal safety approaches are rare. Additionally, only few studies had available geometric
design and built environment data to include in the developed safety performance functions to reveal their impacts on safety. This section provides an overview of the existing literature on exposure measures, geometric design and built environment attributes, integrating safety and mobility in modeling, surrogate safety measures as well as studies using GPS data focusing mainly on cyclist-related research.

1.4.1. Exposure Measures

Safety studies without exposure data (counts) are not able to control for volumes when measuring the effects of contributing factors on injury occurrence and are unable to compute cyclist risk of injury. Exposure measures are among the most important components of safety performance functions and perhaps the biggest determinant of injury occurrence and risk for all modes. Collecting count data for cyclists and motor-vehicles manually, is very time consuming and requires many resources. New technologies have begun to emerge and automated procedures for counting vehicles and cyclists are emerging to complement manual count data. The new automated methods and equipment include, magnetic plates for motor-vehicles, microwave radars, surface or embedded bicycle loop detectors, pneumatic tubes (piezometric), passive and active infrared, Bluetooth detectors, video, Smartphone GPS applications as well as others for cyclists (National Bicycle and Pedestrian Documentation Project, 2009). Some of this technology can distinguish between cyclists and pedestrians, such as active infrared, and some can be set up in one location and then be easily relocated to new locations, such as infrared sensors, pneumatic tubes and video.

Studies which introduced measures of cyclist and motor-vehicle exposure, whether aggregated or disaggregated into different movements (left-turn, right-turn and through movements), arrived at the same conclusion (Elvik, 2009; Jacobsen, 2003). There is a non-linear relationship between cyclist and motor-vehicle flows on injury occurrence (Elvik, 2009). As bicycle flows increase, so does the absolute number of cyclist injuries, however the individual risk faced by each individual cyclist declines. This relationship has been referred to as the "safety in numbers" effect with respect to bicycle flows (Jacobsen, 2003). If the number of cyclists remains constant and there is an increase in motor-vehicle flow, this would cause an increase in cyclist injuries as well as in the risk per individual cyclist. A comprehensive literature review was produced by Elvik (2009) and reports the elasticities obtained in previous studies. Elvik (2009) summarized that when bicycle flows increase by 10%, bicycle injuries are expected to increase by
3% to 6.5% and when motor-vehicle flows increase by 10%, bicycle injuries are expected to increase by 5% to 7.5%. This relationship between traffic volume and safety has endured since Smeed’s Law was developed in 1949 (Smeed, 1949).

A recent study by Nordback et al. (2014) was carried out across intersections in Boulder, Colorado. This study identified that cyclist-vehicle collisions increase non-linearly with both vehicle and cyclist flows. The number of collisions per cyclist decreases as bicycle flows increase and intersections with fewer than 200 cyclists per day have been identified as the most dangerous. Therefore, as expected and in accordance with previous work, the safest intersections are those with high cyclist and low traffic volumes.

Despite the growing literature seeking methods and tools to obtain exposure measures, some gaps have still not been filled. The majority of studies to date have involved a relatively small sample of sites, usually only signalized intersections. For example, Brüde and Larsson (1993) modeled cyclist injury occurrence at fewer than 400 junctions in Sweden with no distinction between the type of junction. In other words, within the small number of sites studied, some were signalized intersections, some non-signalized intersections and some were roundabouts. Wang and Nihan (2004) modeled cyclist injury occurrence at a mere 115 signalized intersections in the Tokyo Metropolitan Area. These studies are examples to show that the majority of cyclist safety studies have been carried out in Europe and Asia, have only considered a small sample of sites and have focused mainly on signalized intersections or have not made a distinction between facility types. Also, previous work has relied on short term (manual) counts which have not been adjusted for the time, day, month and weather conditions when the counts were taken. This work responds to the need for exposure measures at non-signalized intersections and along road segments and is capable of converting short-term counts into average annual daily bicycle flows which can then be mapped in the entire network of intersections and segments. Average annual daily bicycle flows, AADB, is the ideal measure of bicycle exposure since it represents, exactly like its name, the average flow one would expect at that site for the entire day during the year. AADB is obtained by adjusting the short-term counts for each site based on the hour(s), day and month that the counts were taken.

1.4.2. Geometric Design and Built Environment Characteristics

Dumbaugh and Li (2010) emphasized that aside from the effects of traffic exposure, site-specific attributes also affect cyclist injury occurrence. To date, most studies accounting for the effects of
geometric design (GD) and built environment (BE) characteristics have studied cyclist safety from only a small sample of sites.

Overall, cyclist safety studies accounting for the effect of GD and BE have been carried out in Europe and Asia and have identified the following effects:

- In Gothenburg, Sweden, a before-after study was carried out to evaluate the safety effect of raised bicycle crossings (Gårder et al., 1998). Gårder et al. (1998) found that the 44 crossings given this treatment attracted 50% more cyclists in the after period compared to the before period and the safety per cyclist improved by 20%.
- In Copenhagen, Denmark, Jensen (2008) studied the safety effects of painting intersection crossings blue. The case where all four crossings are painted blue resulted in an increase in rear-end collisions as well as right-angle collisions with motor-vehicles, since drivers have placed all their focus on the crossings and not on the traffic signals.
- Oh et al. (2008) revealed that bicycle crashes at urban intersections in Inchon, Korea, increase with increasing average daily traffic volume, number of driveways and in the presence of crosswalks and industrial land use. Crashes were found to decrease with increasing sidewalk widths and in the presence of bus stops and traffic.
- A study in Charlotte, North Carolina, found that cyclists are three to four times more at risk when traveling along segments without any on-street lane compared to there being one (Pulugurtha and Thakur, 2015). Also this study found that providing wider on-street bicycle lanes as well as fewer driveways and intersections along the road, reduce both the number of crashes and cyclist risk.
- Recent studies have been carried out to evaluate the safety effectiveness of cycle tracks (Lusk et al., 2013; Reynolds et al., 2009; Teschke et al., 2012). Overall, cycle tracks were found to have a greater number of cyclists than locations without bicycle infrastructure and the risk of injury was found to be much lower.

In the literature, very little work has been done to thoroughly evaluate the effects that geometric design and built environment characteristics, specific to North America, have on both cyclist injury and bicycle activity outcomes. Furthermore, non-signalized intersections have received much less attention than signalized intersections. To date, the literature that has been interested in studying the effects of GD and BE attributes, have focused on one attribute to evaluate its effect but have not incorporated a wide variety of attributes that can affect cyclist safety and
bicycle activity. In other words, the studies by Gårder et al. (1998), Jensen (2008), Lusk et al. (2013), Reynolds et al. (2009) and Teschke et al. (2012) focused solely on raised bicycle crossings, painted bicycle crossings and cycle tracks respectively. Oh et al. (2008) and Pulugurtha and Thakur (2015) considered more attributes but neither study was carried out in Canada with an extensive dataset of GD and BE attributes to study.

1.4.3. Integral Bicycle Activity and Safety Modeling Approach

Cyclist, pedestrian and motor-vehicle safety studies are all important, but combining all of these into one framework is even more useful and preferable to provide safety for all modes sharing the same facilities. Multimodal approaches are especially needed in locations where there is a high mix of the different modes such as, busy downtown intersections where there are many motor-vehicles, cyclists and pedestrians all trying to reach their destinations. Multimodal safety is a key element in the development of sustainable cities.

Despite its importance, multimodal safety approaches at the city level remain scarce in the literature. The main reason for this is likely due to the unavailability of data and the resources required to collect this data. As a result, most studies have modeled either pedestrian, cyclist or motor-vehicle accident occurrence.

Of the multimodal studies, it is worth mentioning the work of Beck et al. (2007) who compared injury risk between cyclists, pedestrians and vehicles based on daily person-trips in the United States. This study found that motorcyclists, cyclists and pedestrians are 58.3, 2.3 and 1.5 times more likely to be fatally injured, respectively, compared to motor-vehicle occupants.

Pucher and Dijkstra (2003) revealed that there are two issues preventing Americans from switching from motorized to non-motorized modes. These are: walking and biking are perceived as unsafe and these modes are inconvenient and even infeasible. Lessons can be learned from Europe, specifically the Netherlands and Germany, to tackle these two issues. Comparative studies investigating the risk of cycling with respect to other modes remains rare in the literature. Also, the current literature has made minimal effort to use Bayesian methods to identify dangerous intersections for cyclists and pedestrians while also correcting for explanatory variables such as traffic conditions and controls as well as geometric design and built environment characteristics. Modeling frameworks dealing with the endogeneity problem are also rare in the current literature.
1.4.4. Surrogate Safety Measures

Traditional safety analysis models injury occurrence or severity using observed and reported injury data. A new, complementary approach has been growing in the literature which involves identifying a measure that is observable, occurs more frequently than accidents, does not involve any contact or collision between road users and can be related to accidents. The measures described have been named surrogate safety measures since they are alternative ways to represent safety. The main advantages of surrogate safety measures are that there is no need to wait several years to witness sufficient accidents required for analysis and that the safety of sites can be identified and treated prior to witnessing a high number of injuries which means that lives can be saved.

The Federal Highway Administration (FHWA) identifies that there are many surrogate safety measures, including time-to-collision (TTC), post-encroachment-time (PET), gap time (GT) and deceleration rate (DR) (Gettman and Head, 2003). Speed and deceleration distributions can also be used. Specific to intersections, additional surrogate measures include, delay, travel time, approach speed, percent stops, queue length and red-light violations. Also, TTC, PET and DR can be used to measure the severity of conflicts (Gettman and Head, 2003).

To date, most surrogate safety measures have been developed and tested on vehicles but these methods can be applied to cyclists as well. Of the studies carried out for cyclists, we can mention:

- A recent study that computed TTC values to identify the severity of cyclist interactions with vehicles at one busy intersection in Vancouver, British Columbia (Sayed et al., 2013).
- Another recent study computed PET values to evaluate cyclist-vehicle interactions at signalized intersections in Ottawa, Ontario (Kassim et al., 2014).
- Using video data collected at one signalized intersection in China, another study investigated cyclist behaviour, specifically cyclist crossing speed, gap and lag acceptance and group riding behaviour (Ling and Wu, 2004).

Some studies have used video data and surrogate safety measures to evaluate safety treatments for cyclists.

- Bicycle boxes have been implemented in North American cities including Portland, Oregon and Montreal, Quebec, and provide an advanced stop area where cyclists can wait for the light to turn green ahead of the vehicles, improving their visibility to motorized traffic. Dill et al. (2012) and Zangenehpour et al. (2015a) used video data collected before and after this intersection
treatment to study the safety effects of bicycle boxes. Overall these works identified that conflicts between cyclists and vehicles decrease in the presence of bicycle boxes.

- A very recent study in Montreal studied the safety effectiveness of cycle tracks using an automated method to obtain PET as a surrogate measure of the interactions between cyclists and turning vehicles traveling in the same direction (Zangenehpour et al., 2015b).

Some interesting studies which were carried out for vehicles thus far, include:

- In Korea, Lim et al. (2002) developed a methodology to detect red-light, speed, stop line and lane violations.
- A study in Boca Raton, Florida proposed an approach to account for safety at intersections based on surrogate safety measures, to optimize the traffic signal phasing with respect to both safety and efficiency (Stevanovic et al., 2013).

The majority of studies looking at TTC and PET, are based on video data and the values are computed either manually or automatically. Using GPS data, most studies focus on speed and acceleration profiles, as well as delays, since these are fairly easy to extract from this source of data. According to Balasha et al. (1980), for vehicles, if a vehicle applies the brakes, this is a sign of conflict and can be used as a surrogate safety measure. Hard braking is an evasive action which had it not been done, would have resulted in an accident. This same theory is likely applicable for cyclists and will be explored in this work.

- A study in Atlanta, Georgia, showed that the frequency of hard braking is correlated with crash occurrence for vehicles (Jun, 2006).
- One study using GPS data, looked at the braking behaviour of both vehicles and cyclists travelling along a predefined track (Dang et al., 2014).
- Another study used GPS data to identify design inconsistencies on rural roads and their effects on vehicle speed (Cafiso and Cerni, 2012).
- A naturalistic driving study in Virginia, defined a level of critical jerk (rate of change of acceleration) to identify critical braking situations (Bagdadi, 2013). Bagdadi (2013) used a Poisson regression model to assess the relationship between the number of critical braking events and the number of crashes.

While surrogate safety measures are important in order to provide a proactive approach to safety, research is still needed to ensure that these surrogate measures can in fact be related to accidents. There is extensive literature that has tackled the link between conflicts and behaviour.
with accidents for vehicles (Parker et al., 1995). A study by Parker et al. (1995) found the there exists a relationship between jerk rate and drivers’ self-reported involvement in accidents.

While studies have set out to correlate the above conflict measures, no attempts have been made to correlate the intersection measures with actual accident data, rather these measures are simply understood and interpreted as being more dangerous such as long delays (Gettman and Head, 2003).

According to Williams (1981), evidence exists that conflicts are similar to accidents and have the same characteristics and only differ in that conflicts are lacking any physical impact. Zeeger highlights that accident and conflict data complement one another (1977).

Many studies have emerged in the literature seeking to develop surrogate safety measures to identify dangerous locations prior to witnessing many accidents. Furthermore, understanding the relationship between driving behaviour and conflicts with accidents is important for safety.

The development of surrogate safety measures for cyclists, other than conflicts with vehicles at intersections, has yet to be thoroughly addressed in the literature. In this thesis, the Smartphone GPS data makes it possible to extract the number of dangerous decelerations that occurred along all segments and through all intersections in the network which can then be tested to see how correlated this measure is with observed accidents at the same locations. Smartphone GPS data for cyclists is a very new source of data for cyclists and therefore the development of surrogate safety measures from this source and for the entire network, have yet to be developed in the literature. However, being able to identify where cyclists are most at risk at the network level, as is carried out in this thesis, without having to witness any accidents, has great implications for preventing injuries and fatalities.

1.4.5. Cyclist GPS Data

To date, studies have begun to extract level-of-service measures using GPS units, cellular data and Smartphones.

- One study used cell phone call records to generate origin-destination (OD) matrices for vehicles traveling in the city of Dhaka, Bangladesh (Iqbal et al., 2014).
- Another study in the United States recorded locations every time a cell phone connected to the cellular network, by placing a call, sending a text message or connecting to the internet (Calabrese et al., 2011).
• A study in Israel, used cell phone data to obtain speeds and travel times for vehicles (Bar-Gera, 2007).
• A study in India, equipped personnel with cell phones who then rode inside the public transit buses to continuously collect location data (Satyakumar et al., 2014). This data was used to predict bus arrival times.
• A study in Union City, California used Smartphone GPS data to process position and speed data for vehicles in real-time for traffic monitoring (Herrera et al., 2010).
• In the City of Minneapolis, Minnesota, El-Geneidy et al. (2007) equipped cyclists with GPS units to obtain speed data to develop a model to predict cyclist speeds along off-street facilities, on-street bicycle lanes, and on-streets without any facilities. Cyclists were found to ride faster along on-street facilities.
• In Leeds, United Kingdom, Parkin and Rotheram (2010) equipped 17 cyclists with GPS units to obtain speeds on level terrain as well as when cycling uphill and downhill to obtain a value for cyclist free-flow speeds for each case. This study found that cyclist speed when traveling uphill is more affected than speed when traveling downhill.

To date, cyclist GPS data from a Smartphone application has not been used to extract the number of cyclists and trips which serves as vital input in safety analysis, infrastructure management, planning and development, mobility analysis as well as many others. Furthermore, GPS data has only focused on level-of-service measures and has not been used in safety studies. Also, the main benefit of this source of cyclist data is that since data is collected throughout the entire network, all intersections and road segments can be used to study and map cyclist safety and mobility measures. This data has great implications for network screening since the entire population of sites are compared to one another as opposed to using other sources of data which are point-specific and unable to be used for network screening.

1.5. ORIGINAL CONTRIBUTIONS

This research is unique since it addresses the shortcomings in the existing research by responding to the need for computing and mapping cyclist activity and injury risk throughout an entire urban network of intersections and road segments. This work also addresses the importance of multimodal approaches for safety by proposing a multimodal Bayesian framework. This work incorporates all levels from data collection, monitoring, integrating to processing and analysis in
order to create decision making tools from which informative and appropriate decisions can be made. This is a continuous process which can constantly be updated with the availability of new data and the development of new analysis techniques.

This research is the first of its kind to integrate a wide variety of data, either collected by the McGill University Transportation Group, or provided by organizations we are collaborating closely with, such as the Montreal Department of Public Health, Urgences-Santé, the city of Montreal and others. The majority of safety studies currently in the literature suffer due to the lack of disaggregate exposure measures covering a large sample of intersections over many hours or adjusted using expansion factors. Also, the majority of safety studies do not have such a rich dataset of geometric design and built environment attributes for such a large sample of sites. Also, this work is one of the first to use Smartphone GPS data for cyclists for safety and mobility purposes. Using this data, this study will be able to map not only injuries but also activity levels and risk for cyclists throughout the island of Montreal.

Controlling for the level of cyclist and vehicle flows and including GD and BE attributes are important inputs towards the development of safety performance functions for all modes along all types of facilities and infrastructure. Knowledge of their effects serve as inputs into identifying hotspots and identifying the appropriate treatments to provide safer facilities. Dumbaugh and Li (2010) sought to demonstrate that driver error is not random, rather it is a result of the built environment. And further that it is the built environment that plays a role in traffic-related crashes for all modes in an urban environment. This emphasizes that in addition to accounting for levels of traffic flow, geometric design of roads and intersections are also important factors affecting injury occurrence for all modes.

1.6. GENERAL METHODOLOGY

This research is supported by a conceptual framework, shown in Figure 1-1 and is guided by the need for a tool to link exposure data collection procedures to micro-level geometric design and built environment characteristics attributed to the entire population of sites. This information can then be used to analyze safety and mobility from a multimodal perspective guiding decision making tools and leading to safety and mobility improvements. This framework has three main sections, represented by the dotted lines in Figure 1-1, and are: 1) data collection, monitoring and integration, 2) general model components and analysis, and 3) decision making tools. This framework can be
applied at any period of time and to any urban area. Each section builds upon one another and forms a continuous cycle. The capability to draw conclusions leading to decision making which comes from good analysis, begins with and depends on the availability of reliable data, leading to the first section: data collection, monitoring and integration. Emerging counting technology is capable of collecting data for long periods of time, not only counting for different hours of the day as are done with manual counts but also, if left in place for extended periods of time, counts can also be obtained for different days of the week and months. By combining the data obtained from all the counters, spatial trends per hour, day and month can also be extracted. This data will enable studies to consider both temporal and spatial trends in cyclist, pedestrian and motor-vehicle flows allowing for disaggregate measures of exposure. Having these adjusted flows brings us to section 2: general model components and analysis. The exposure data adds to the inventory of intersections which has geometric design, built environment as well as traffic control data specific to a large sample of signalized and non-signalized intersections on the island of Montreal. This rich dataset permits studies to account for many factors contributing to the safety and level-of-service for all modes, which represents the last section: decision making tools. With knowledge of the safety and mobility measures for all modes using each facility, decisions can be made as to where and how to allocate funds towards improvements following network screening. Due to the importance of providing safety and adequate level-of-service for all road users within a limited budget, decision making tools are necessary for budget allocation in the most effective way as to maximize safety and mobility benefits. All the data and analysis results can also be used to map risk and traffic intensity on the island, carry out corridor analysis and evaluate the effectiveness of countermeasures for safety and level-of-service improvements. In the simplest way, this research involves collecting, processing and then analyzing data to be able to identify hazards in terms of safety and mobility. This study begins by installing both permanent and semi-permanent counting equipment combined with manual traffic counts. This data can be combined with speed and other flow attributes to form a database from which many different studies ranging from safety to facility design can be carried out and used to support the process of decision making.

In this work, the framework is applied to cyclists on the island of Montreal. The three main sections are applied as follows:

**Section 1:** Data Collection, Monitoring and Integration: Since at least 2008, automatic bicycle counters have been installed in specific locations on the island of Montreal. The majority of the
counters collected data from a site for as little as two months to as many as five years of continuous count data. This data was then processed to extract hourly, daily and monthly expansion factors. Weather conditions were also downloaded and linked to each hour of counts to determine their effects on cyclist numbers to adjust the counts accordingly. In the meantime, manual counts were carried out at a large sample of signalized and non-signalized intersections. The manual and automatic counts can then be integrated to convert the short-term counts into average annual daily bicycle values (AADB), the preferred unit of exposure in safety analyses. Very recently, a new source of automatic cyclist counts was developed for the city of Montreal, a Smartphone GPS application, Mon RésoVélo. When a cyclist turns on the application and starts their trip, the application collects second-by-second locations and timestamps, using the Smartphone's built-in GPS, for their entire route from origin to destination. Once again, combining the counts from the application with the expansion factors provides a new approach to compute AADB in the network. AADB based on the GPS data can then be correlated with AADB based on the manual counts to check the accuracy of the GPS data which has a wide range of safety and mobility applications that benefit from being able to consider the entire population of sites in the urban network. The automatic counters which are still installed, as well as the Smartphone application, continue to collect data and can be monitored to observe the change in trends from year to year, to update the analysis and to be integrated into new analyses.

Section 2: The main output of section 1 is AADB which, as mentioned, is the preferred unit of exposure in safety analyses. In addition to flows, geometric design and built environment characteristics are also important inputs in safety and mobility analyses. Geometric design attributes incorporate traffic control attributes and bicycle infrastructure whereas built environment attributes include land use mix, public transit supply and network connectivity. Developing safety performance functions and studying mobility measures are carried out by combining all the above mentioned data.

Section 3: The compilation of all this data enables a decision making tool, which ideally should include all modes but in this case provides a complete view of one mode including safety and mobility outcomes in the entire network. At this stage, sites can be ranked in terms of injuries and risk in order to identify the most dangerous locations and carry out network screening to determine sites in need of safety improvements. Since all sites are being compared, this provides a complete comparison which can ensure that the most dangerous sites have in-fact been selected for
treatments, which as a result enables good budget allocation and most importantly, results in the greatest safety improvements in the network. Additionally bicycle flows can be mapped and used to identify if and where bicycle infrastructure is lacking, by clearly showing areas on the island of Montreal where there is a high number of cyclists but no infrastructure. More importantly, cyclist risk in the network can also be mapped and by merging this result with bicycle flows, this can identify areas where a high number of cyclists are exposed to a high risk and propose solutions. For mobility applications, speeds and delays can also be mapped in the network and used to identify segments with low cyclist speeds and intersections with high delay for cyclists to propose level-of-service (LOS) improvements. This information can also be combined with safety measures to develop a navigation tool for cyclists to minimize their risk and/or travel time.
1.7. ORGANIZATION OF THE DOCUMENT

This thesis is organized into seven chapters. The first chapter contains the introduction and provides background, context, objectives, a review of the literature, original contributions and the general methodology. This thesis is manuscript-based so chapters two to six contain five separate journal papers for which the author is the primary author. Chapters two, three and four have already been
published in a peer-reviewed journal, Accident Analysis and Prevention. Chapter five has been submitted to a peer-reviewed journal, Transportation Research Part F: Traffic Psychology and Behaviour: Special Issue on Road Safety as Reflected by Empirical Non-crash Data. Chapter six will be submitted to a peer-reviewed journal, Transportation Research Part C: Emerging Technologies.

1.8. REFERENCES


Chapter 2

Cyclist Activity and Injury Risk Analysis at Signalized Intersections: A Bayesian Modeling Approach
Chapter 2: Cyclist Activity and Injury Risk Analysis at Signalized Intersections: A Bayesian Modeling Approach

Jillian Strauss\textsuperscript{a}, Luis Miranda-Moreno\textsuperscript{a} and Patrick Morency\textsuperscript{b}

\textsuperscript{a}Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, Canada
\textsuperscript{b}Montreal Department of Public Health, Montreal Health and Social Service, Montreal, Canada

Presentation at the Transportation Research Board 92\textsuperscript{nd} annual meeting, in Washington D.C, 2013 and has been published in Accident Analysis and Prevention, 2013.

2.1. ABSTRACT

This study proposes a two-equation Bayesian modeling approach to simultaneously study cyclist injury occurrence and bicycle activity at signalized intersections as joint outcomes. This approach deals with the potential presence of endogeneity and unobserved heterogeneities and is used to identify contributing factors associated with both cyclist injuries and volumes. Its applicability in the identification of corridors at high-risk is also illustrated. An extensive inventory of a large sample of signalized intersections on the island of Montreal is used as the application environment. This inventory contains not only disaggregate vehicular traffic volumes and bicycle flows but also geometric design, traffic controls and built environment characteristics in the vicinity of the intersections. Among other results, this study identifies the importance of both bicycle and motor-vehicle flows on cyclist injury occurrence and further emphasizes the importance of turning motor-vehicle movements. It was also found that the presence of bus stops and total crosswalk width have a positive effect on cyclist injury occurrence whereas the presence of a raised median has a negative effect. The results also reveal that bicycle activity through intersections increases as employment, number of metro stations, land use mix, area of commercial land use type, length of bicycle facilities increase as well as in the presence of schools measured between 50 and 800 metres from the intersection. Also, intersections with three approaches are expected to have fewer cyclists than intersections with four approaches. Using Bayesian analysis, expected injuries and injury rates
were computed and used to rank corridors. We found that corridors with high bicycle volumes, located mainly in the central neighbourhoods of Montreal, have lower risk despite having a high number of cyclists riding them along each day. Most importantly, since there are more cyclists, these corridors have a greater chance of ranking high therefore confirming the "safety in numbers" effect.

2.2. INTRODUCTION

While cyclist numbers continue to rise and the benefits continue to be enjoyed, cycling in urban environments still comes with serious safety concerns, in particular at intersections. An intersection is a complex area where many interactions can occur between cyclists, motor-vehicles and pedestrians. It is then not surprising to observe that an important proportion of cyclist injuries occur at intersections. For instance, from 1999 to 2003, an average of 950 cyclists were injured every year on the island of Montreal, 58% of which occurred at an intersection (Morency and Cloutier, 2005; Morency et al., 2012). Also, on average 3 cyclists are killed per year in this territory. This issue becomes more relevant in the current context in which bicycle usage and infrastructure are growing in North American cities including Montreal.

Given the importance of this issue, cyclist injury occurrence at intersections is a topic that has started to receive the attention it deserves. Various studies have investigated the link between motor-vehicle and bicycle flows on crash frequency at intersections (Brüde and Larsson, 1993; Miranda-Moreno et al., 2011; Wang and Nihan, 2004) while very few have focused on analyzing the safety effectiveness of bicycle facilities (Lusk et al., 2011). Despite these important efforts, gaps in the current literature still exist and little is known about the effects of traffic control, geometric design and built environment factors on cyclist injury risk (Strauss and Miranda-Moreno, 2012). To our knowledge, no previous studies have tried to model cyclist injury risk and bicycle activity simultaneously. Observed geometric design and built environmental attributes as well as unobserved factors can be associated with both outcomes. For example, the implementation of an intersection treatment (e.g., installation of a bicycle facility, a change in intersection geometry or traffic controls) may be associated with an increase or decrease in bicycle volumes as well as in the number of injuries. Due to observed and unobserved factors that can potentially affect both cyclist injuries and bicycle flows, there is a need to model these two outcomes simultaneously. This is often referred to as the endogenous problem. When identifying contributing risk factors or
identifying locations at high risk, this statistical issue should be taken into account. An appropriate method will also help to provide better risk estimates and identify factors associated with injury occurrence.

This study has two main objectives, to: i) propose a two-equation Bayesian model to study cyclist injury occurrence and bicycle activity simultaneously, and ii) study the role of geometric design, built environment and traffic control characteristics on injury occurrence and bicycle activity. The proposed approach is then used to rank bicycle corridors based on the expected number of injuries and injury rates.

This paper is broken down into several sections. Section 2 offers a literature review. Section 3 describes the methodology followed by the site selection and data in Section 4. Section 5 discusses the results and Section 6 presents the conclusions drawn from this study as well as the limitations and directions for future work.

2.3. LITERATURE REVIEW

Cyclist safety studies examining the link between motor-vehicle and bicycle volumes on crash occurrence at intersections have been published. A summary of these studies has been reported by (Elvik, 2009). Overall, previous studies have identified that there is a non-linear relationship between both bicycle and motor-vehicle flows on cyclist injury occurrence. The non-linear relationship between bicycle flows and injury rates has been referred to as the “safety in numbers” effect, which means that as bicycle flows increase, so does the absolute number of cyclist injuries, however the individual risk (rate) faced by each cyclist declines (Jacobsen, 2003). Overall, past studies have found that the frequency of bicycle accidents increases by 3% to 6.5% when bicycle flows increase by 10% (Elvik, 2009).

Few studies have looked at the effects of geometric design and built environment characteristics on cyclist injury outcomes at intersections. Among these studies, we can refer to Gårder et al. (1998), Jensen (2008), Oh et al. (2008) and Ma et al. (2010), which have all been carried out in Europe or Asia. In Sweden, Gårder et al. (1998) found that raised bicycle crossings reduced motor-vehicle speed which as a result reduced the number of accidents occurring between motor-vehicles and cyclists. The authors emphasize, however, that accident risk is more sensitive to bicycle speeds than to motor-vehicle speeds which were found to increase after the implementation of the raised crossing. In Copenhagen, Denmark, Jensen (2008) studied the safety
effects of painting intersection crossings blue. The case where all four crossings are painted blue resulted in an increase in rear-end collisions as well as right-angle collisions between motor-vehicles, since drivers have placed all their focus on the crossings and not on the traffic signals. Oh et al. (2008) revealed that bicycle crashes at urban intersections in Inchon, Korea, increase with increasing average daily traffic volume, number of driveways and in the presence of crosswalks and industrial land use. Crashes were found to decrease with increasing sidewalk widths and in the presence of bus stops and traffic calming measures. In Beijing, China, Ma et al. (2010) found that arterials with many lanes of traffic, high levels of motor-vehicle flow and high speeds present greater risk of severe injury for cyclists. Bicycle activity is a key variable when modeling and estimating cyclist injury risk at intersections. Recent studies have looked at the factors associated with bicycle activity through intersections. Three recent studies in California as well as one even more recent study in Montreal have investigated the link between bicycle activity and geometric design, built environment factors, road and transit characteristics as well as socio-demographic attributes (Griswold et al., 2011; Haynes and Andrzejewski, 2010; Jones et al., 2010; Strauss and Miranda-Moreno, 2011). Using regression analysis, these studies have found that employment, presence of schools, metro stations, bus stops, land use mix, commercial retail properties, proximity to a major university, mean income, bus frequency, presence and proximity of bicycle facilities, road network connectivity as well as non-hilly terrain have a positive effect on bicycle activity, whereas average street length and the presence of parking entrances/exits have a negative effect on bicycle activity.

Despite the importance of previous studies, some research gaps still exist. In the emerging literature, cyclist injury occurrence and bicycle activity have been modelled independently. This may prevent us from understanding the direct and indirect effects of observed and unobserved factors on these two outcomes (Miranda-Moreno et al., 2011). For instance, some geometric attributes (e.g., presence of bicycle facilities) can have an indirect effect on cyclist safety through an increase or decrease in bicycle activity. Very few studies, if any, have accounted for endogenous effects which may arise since the relationship between cyclist injury occurrence, bicycle flows and associated factors, for which the direction of causality may be unknown or for which there are unobserved factors, can cause endogeneity. When endogenous effects are present and left unaccounted for, this can lead to biased and incorrect parameter estimates, however, accounting for endogenous effects often requires more complex estimation procedures (Lee, 2007). In
addition, cyclist safety studies at intersections are rare in North America and most have been carried out in European and Asian cities (Nordback et al., 2012). While a few studies have been carried out in the United States and Canada, these have mainly focused on cyclist injuries at the bicycle facility, city or town level and did not focus on intersections (junctions) as the unit of study (Miranda-Moreno et al., 2011). Also, most of these studies have used total cyclist and total motor-vehicle flows as a measure of risk exposure and have not considered more disaggregate measures to allow a complete observation of the impact of the different movements. These studies have also not considered geometric design, built environment and traffic control characteristics related to the layout and location of the intersections. Furthermore, previous studies are based on old data and relatively small sample sizes. Finally, despite the extensive literature on hotspot identification, little effort has been done to map cyclist risk using Bayesian methods for the identification of dangerous corridors while also correcting for explanatory variables (traffic conditions, geometric design and built environment characteristics).

This research tries to address some of the above mentioned research gaps by developing a methodology that includes the combination of different sources of data and the use of disaggregate motor-vehicle flows to simultaneously investigate observed and unobserved factors associated with cyclist injury occurrence and bicycle activity.

2.4. METHODOLOGY

A two-equation simultaneous Bayesian framework is proposed to study cyclist injury occurrence and bicycle activity (also referred to as volumes or flows). In the model, cyclist injuries are a function of bicycle flows, motor-vehicle flows, geometric design, traffic control and built environment characteristics as well as some unmeasured factors (unobserved site-specific characteristics, cyclist behaviour, etc.). Bicycle flows can be a function of the built environment, geometric design variables, weather conditions and other observed and unobserved variables. In our previous work, these two outcomes were studied independently (Miranda-Moreno et al., 2011; Strauss and Miranda-Moreno, 2011; Strauss and Miranda-Moreno, 2012 ). However, unobserved factors may be affecting both outcomes and potentially causing an endogeneity problem. The two outcomes can be represented as shown in Equation 1:

\[ Y_i = f(t_i, Z_i, V_i, x_i, \varepsilon_{1i}) \] (1)

\[ Z_i = f(V_i, w_i, \varepsilon_{2i}) \]
where \( Y_i \) = number of cyclist injuries observed at a given intersection \( i \) \((i = 1, \ldots, n\) intersections) during a period of time \( t_i \), \( Z_i \) = average annual daily bicycle flow at the same intersection calculated based on 8-hour counts which have been expanded, \( V_i \) = average annual daily motor-vehicle traffic flow at the same intersection, disaggregated into left turn, right turn and through movements entering the intersection in all approaches, \( x_i \) = vector of variables including geometric design and built environment characteristics (population density, presence of metro (subway) or bus stops, bicycle facilities in the vicinity of the intersection, etc.) as well as traffic controls (left turn signals, road and crosswalk lengths, presence of median, pedestrian/bicycle half or full phase, etc.), \( w_i \) = vector of variables including mainly built environment characteristics (population density, presence of metro or bus stops, bicycle facilities in the vicinity of the intersection, etc.) as well as some geometric design and traffic controls. Note that \( w_i \) and \( x_i \) can contain the same or different variables, \( t_i \) = time period of observation in years. This varies across intersections since for intersections with bicycle facilities, only the injuries that occurred after its installation are included and \( \varepsilon_1 \) and \( \varepsilon_2 \) = correlated error terms representing unobserved factors influencing both injuries and flows. Some unmeasured design and comfort (level-of-service) conditions might affect both injury occurrence and bicycle activity.

To model these two outcomes simultaneously, a bivariate mixed Poisson model with correlated Lognormal error terms is proposed. This is formulated as:

\[
\begin{align*}
Y_i \mid \theta_{iY} & \sim \text{Poisson}(t_i \theta_{iY}) \\
Z_i \mid \theta_{iZ} & \sim \text{Poisson}(\theta_{iZ}) \\
\text{With} & \\
\theta_{iY} & = V_i^{\alpha_1} Z_i^{\alpha_2} \exp(\beta_0 + \beta_1 \cdot x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_{iY}) \\
\theta_{iZ} & = \exp(\gamma_0 + \gamma_1 \cdot w_{i1} + \ldots + \gamma_k w_{ik} + \varepsilon_{iZ}) \\
\varepsilon_i \mid T & \sim N_M(0, T),
\end{align*}
\]

where \( N_M() \) stands for the bivariate normal distribution with mean vector 0 and covariance matrix \( T \). In addition, \( \alpha_1, \alpha_2, \beta = (\beta_0, \ldots, \beta_k) \) and \( \gamma = (\gamma_0, \ldots, \gamma_k) \) are model regression parameters to be estimated from the data. Priors on these parameters are assumed to follow a non-informative or informative Gaussian (normal) distribution. Non-informative priors are assumed to have a mean equal to 0 and a very large variance, e.g., \( \beta_i \sim N(0,1000) \). Informative priors can be assumed for some parameters with prior knowledge. Also, the regression parameters are mutually independent.
Once the best model is identified in terms of goodness of fit and parameter credible intervals, the posterior analysis can be carried out to identify hotspots at different spatial levels. This task requires good estimates in order to obtain accurate and comparable risk measures. Bayesian analysis is very popular in road safety literature. The advantage of Bayesian analysis is that posterior risk estimates can be computed to identify intersections, corridors or areas at high risk. Popular posterior risk measures used in safety include the posterior mean number of injuries, the posterior injury rate, the potential of risk reduction, the posterior distribution of ranks and the posterior probability that a site is the worst (Miaou and Song, 2005; Miranda-Moreno, 2006; Schlüter, 1997). Among these, the posterior mean of injury frequency at intersection \( i \) (denoted by \( \bar{\theta}_{iY} \)) is one of the most popular injury risk measures and is computed according to Equation 2.

From this measure, the posterior injury rate per million cyclists per unit of time – denoted by \( \bar{R}_{iY} \) – can be computed as shown in Equation 3:

\[
\bar{\theta}_{iY} = E[\theta_{iY} | y_i] = \int_{\theta_{iY}} \theta_{iY} p(\theta_{iY} | y_i) d\theta_{iY} \tag{2}
\]

\[
\bar{R}_{iY} = \frac{\bar{\theta}_{iY} \times 10^6}{365 \cdot t_i \cdot Z_i} \tag{3}
\]

As previously defined, \( Z_i \) represents the average annual daily bicycle flow crossing the intersection. To estimate the injury rates at the corridor level, an average corridor value is obtained according to \( \bar{R}_Y = \frac{n_c}{\sum_{i=k}^{n_c} \bar{R}_{iY}} / n_c \) where \( n_c \) stands for the number of intersections.

For each model formulation, Gibbs sampling is used for posterior inference. This is implemented using the open-source software OpenBUGS (Bayesian inference using Gibbs sampling). In order to obtain parameter estimates (posterior means, standard deviations and credible intervals of regression parameters) and the DIC (Deviance Information Criterion) value, 150,000 updates are burned and then, 10,000 additional samples are drawn. For model specification, correlation among geometric design, built environment and traffic control variables was first verified using a correlation matrix to avoid co-linearity issues. Also, a standard Negative Binomial regression model was estimated using a maximum likelihood approach for each of the outcomes to pre-select potential variables for the Bayesian model.
2.5. DATA

A sample of 647 signalized intersections on the island of Montreal, Quebec, Canada is used as the application environment. The study intersections meet the following four conditions: 1) recent cyclist and motor-vehicle counts are available, 2) counts were carried out during the cycling season, between April 1st and November 30th when seasonal bicycle facilities are open, 3) data regarding a large variety of geometric design and built environment characteristics are also available, and 4) the completion dates are available for bicycle facilities starting, ending or passing through the intersections. Figure 2-1 shows a map of the island of Montreal and the intersections studied. This figure identifies that most of the intersections are located in the central neighbourhoods of the island and clearly identifies that some areas of the island are either over or under-represented.

Figure 2-1 Study intersections and average observed injury rates

This research uses an inventory of intersections which combines different types of data and sources such as: a) 8-hour bicycle and motor-vehicle counts collected recently by the Montreal Department of Transportation, b) cyclist injury data provided by the Montreal Department of Public Health (Urgences-santé), c) geometric design and traffic control data collected by our McGill team, as well as d) built environment data from Statistics Canada, DMTI Spatial Inc., STM (Société de
transport de Montréal) and AMT (Agence métropolitaine de transport). The motor-vehicle and bicycle flows, cyclist injury, geometric design, traffic control and built environment data is described below and some summary statistics are provided in Table 2-1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyclist injury count (observed)</td>
<td>0.63</td>
<td>1.32</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Period of observation (years)</td>
<td>5.7</td>
<td>1.25</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Bicycle flows</td>
<td>444.5</td>
<td>714.9</td>
<td>1.66</td>
<td>6433</td>
</tr>
<tr>
<td>Motor-vehicle right turn flows</td>
<td>2586.1</td>
<td>2638.6</td>
<td>0</td>
<td>23843</td>
</tr>
<tr>
<td>Motor-vehicle left turn flows</td>
<td>2653.8</td>
<td>2692.6</td>
<td>0</td>
<td>23792</td>
</tr>
<tr>
<td>Motor-vehicle through flows</td>
<td>19383.9</td>
<td>11088.5</td>
<td>1790</td>
<td>76525</td>
</tr>
<tr>
<td>Presence of bus stops</td>
<td>0.71</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total crosswalk width (sum for all approaches)</td>
<td>68.5</td>
<td>24.1</td>
<td>9.29</td>
<td>245</td>
</tr>
<tr>
<td>Presence of raised median</td>
<td>0.47</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Pedestrian signal (with or without countdown)</td>
<td>0.55</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total number of lanes (sum for all approaches)</td>
<td>6.9</td>
<td>2.6</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>400m Employment ('000)'</td>
<td>1.5</td>
<td>0.97</td>
<td>0.06</td>
<td>4.7</td>
</tr>
<tr>
<td>400m Presence of schools</td>
<td>0.56</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>800m Metro (subway) stations</td>
<td>1.5</td>
<td>1.8</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>800m Land use mix</td>
<td>0.67</td>
<td>0.15</td>
<td>0</td>
<td>0.92</td>
</tr>
<tr>
<td>50m Area of commercial land use ('000)'</td>
<td>0.89</td>
<td>1.4</td>
<td>0</td>
<td>7.8</td>
</tr>
<tr>
<td>800m Length of bicycle facilities</td>
<td>2.1</td>
<td>1.7</td>
<td>0</td>
<td>7.9</td>
</tr>
<tr>
<td>Presence of bicycle lane</td>
<td>0.044</td>
<td>0.206</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Presence of cycle track</td>
<td>0.04141</td>
<td>0.1994</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Presence of three approaches</td>
<td>0.21</td>
<td>0.41</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

1-3 - These variables were measured within a 1) 50 m, 2) 400 m or 3) 800 m radius around the intersections.

With respect to the traffic flow conditions, the Montreal Department of Transportation manually collected both bicycle and motor-vehicle counts in 2009. These counts were carried out during three periods of the day, morning peak (6:00 a.m.–9:00 a.m.), lunch period (11:00 a.m.–1:00 p.m.) and evening peak (3:30 p.m.–6:30 p.m.), providing a total of eight hours of flow data. This dataset was filtered to only include the months for which Montreal's bicycle facilities are open, from April 1st to November 30th.

The city of Montreal has also provided bicycle count data collected from automatic bicycle counters (loop detectors) located along five bicycle facilities in Montreal's central neighbourhoods. Using the automatic bicycle count data, hourly, weekly and monthly expansion factors were obtained (Miranda-Moreno and Nosal, 2011; Strauss, 2012a). As identified in the literature, bicycle flows are sensitive to weather conditions (Miranda-Moreno and Nosal, 2011; Nankervis, 1999).
By joining the automatic bicycle count data with the weather conditions present during each hour of counts, weather models specific to Montreal were developed (Miranda-Moreno and Nosal, 2011). These weather models along with the expansion factors were then applied to convert the 8-hour counts into average annual daily values (Strauss and Miranda-Moreno, 2011). The 8-hour motor-vehicle flows have also been adjusted using expansion factors provided by the city.

This study focuses on cyclist injuries having occurred at the 647 intersections of interest over the period from January 1st, 2003 to July 31st, 2008 (almost six years). Accidents are considered as having occurred at an intersection if they are within 15 m of the centre point of the intersection. This study uses ambulance data instead of police report data since this data has less under-reporting and misallocation problems (Langley et al., 2003). Although ambulance data may be biased towards more severe injuries, in Montreal, this source of data identified more cyclist injuries than police reports. Figure 2-1 shows all 647 intersections, represented by each dot, and the average observed injury rates, represented by the size of the dot. This figure identifies that overall, the intersections located in the central neighbourhoods of the Island have witnessed a greater number of cyclist injuries. Bicycle injury data is provided at the level of the individual and not at the level of the crash. If a crash should involve two cyclists, this would actually be considered as two separate injuries; however this situation arises in a very small percentage of the cases reported in this dataset.

Some intersections have bicycle facilities, either bicycle lanes or cycle tracks that were installed during the six-year study period for which we have injury data. In this case, we are only interested in accidents having occurred after the facility was completed. Therefore we only consider the number of years with injury data, based on the completion date of the facilities. This is accounted for in the cyclist injury model as defined by $t_i$ in Equation 1.

To obtain the geometry characteristics of each intersection, an important data collection campaign was undertaken at McGill University during the summer and fall of 2010 and 2011. Our team collected a rich inventory at almost 80% of signalized intersections in Montreal which includes a wide variety of geometric design and traffic control characteristics such as: number of approaches, number of lanes, type of traffic signals, pedestrian/cyclist phasing, left turn lanes and phasing, presence and width of medians, presence and type of bicycle facilities, crosswalk length and so on. Data collection teams visited each intersection with specific data collection sheets and an odometer.
Built environment characteristics such as land use, urban form and bicycle facility characteristics have been provided from different sources: Statistics Canada, DMTI Spatial Inc., STM and AMT. This data includes population, employment, income, land use, presence of metro (subway) stations, bus stops, street typology, other demographics and road and transit characteristics. These variables were extracted for four different buffer dimensions: 50 m, 150 m, 400 m and 800 m to evaluate the impact of these variables at different distances from the intersection. Although bicycle activity may be better predicted using larger buffer sizes, caution must be taken when selecting variables for the model since the proportion of correlated variables is very likely to increase with increasing buffer sizes.

2.6. RESULTS

The best regression outcomes, selected based on the best DIC values and credible intervals, are presented in Table 2-2. It is important to mention that only variables significant to the 95% level (regression parameters with 95% credible intervals that do not include zero) were retained in the final models. Elasticities were also computed for each variable. It can be shown that the elasticity, \( E = (dy/dx)(x/y) \) for log-linear variables, the flow variables, is equal to the parameter, e.g., \( E = \alpha_1 \) for \( V_i \). For continuous variables, such as crosswalk length, the elasticity is equal to the regression parameter times the mean value of the variable \( E = \beta_j \bar{x}_i \). For dummy variables, the elasticity is computed as \( E = \exp(\beta_j - 1)/\exp(\beta_j) \). Elasticities were evaluated at the mean values (Washington et al., 2003).

Additionally, a sensitivity analysis was carried out to explore the impact of setting alternative priors for the model parameters, e.g., smaller variances in the regression parameters, e.g., \( \beta_j \sim N(0,100) \). The literature has reported regression parameters for bicycle flow that range from 0.3 to 0.65 with a mean value of 0.5 (Elvik, 2009). Accordingly, semi-informative priors were built for the bicycle flow regression parameter, e.g., \( \alpha_1 \sim N(0.5,100) \). From the sensitivity analysis, it was observed that the model results were consistent across different prior specifications. This is not surprising given the fact that the mean number of observed injuries is low but the sample size is relatively large (Miranda-Moreno et al., 2013).
## Table 2-2 Results

<table>
<thead>
<tr>
<th>Cyclist Injury Model</th>
<th>Coefficient</th>
<th>Standard Deviation</th>
<th>95% Credible Interval</th>
<th>Elasticity&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln bicycle flows&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.869</td>
<td>0.071</td>
<td>0.765 - 1.013</td>
<td>8.69</td>
</tr>
<tr>
<td>Ln motor-vehicle right turn flows&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.240</td>
<td>0.039</td>
<td>0.153 - 0.307</td>
<td>2.40</td>
</tr>
<tr>
<td>Ln motor-vehicle left turn flows&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.185</td>
<td>0.050</td>
<td>0.106 - 0.279</td>
<td>1.85</td>
</tr>
<tr>
<td>Presence of bus stops&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.519</td>
<td>0.164</td>
<td>0.196 - 0.842</td>
<td>40.5</td>
</tr>
<tr>
<td>Total crosswalk width</td>
<td>0.009</td>
<td>0.002</td>
<td>0.004 - 0.014</td>
<td>5.84</td>
</tr>
<tr>
<td>Presence of raised median&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-0.351</td>
<td>0.153</td>
<td>-0.640 - -0.037</td>
<td>-42.1</td>
</tr>
<tr>
<td>Constant</td>
<td>-10.08</td>
<td>0.329</td>
<td>-10.66 - -9.58</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bicycle Activity Model</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>400m Employment ('000)</td>
<td>0.285</td>
<td>0.006</td>
<td>0.273 - 0.294</td>
<td>4.39</td>
</tr>
<tr>
<td>400m Presence of schools&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.288</td>
<td>0.016</td>
<td>0.258 - 0.313</td>
<td>25.1</td>
</tr>
<tr>
<td>800m Metro (subway) stations</td>
<td>0.218</td>
<td>0.004</td>
<td>0.213 - 0.226</td>
<td>3.37</td>
</tr>
<tr>
<td>800m Land use mix</td>
<td>0.173</td>
<td>0.015</td>
<td>0.145 - 0.201</td>
<td>1.15</td>
</tr>
<tr>
<td>800m Length of bicycle facilities</td>
<td>0.139</td>
<td>0.005</td>
<td>0.131 - 0.146</td>
<td>2.88</td>
</tr>
<tr>
<td>50m Area of commercial land use ('000)</td>
<td>0.015</td>
<td>0.007</td>
<td>0.005 - 0.029</td>
<td>0.13</td>
</tr>
<tr>
<td>Presence of three approaches&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-0.572</td>
<td>0.022</td>
<td>-0.601 - -0.521</td>
<td>-77.11</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.227</td>
<td>0.006</td>
<td>-0.237 - -0.218</td>
<td></td>
</tr>
</tbody>
</table>

DIC: 7134
Covariance: 1.137

<sup>a</sup>- Ln stands for the natural logarithm
<sup>b</sup>- Elasticities are expressed in terms of a 1% change in the independent variable or a 0 to 1 change in the case of a dummy variable.
<sup>c</sup>- Dummy variables.

### 2.6.1. Factors Associated with Cyclist Injury Risk

Consistent with the literature, these results reveal the important effects of bicycle and motor-vehicle flows on cyclist injury occurrence. According to the elasticities reported in the last column of Table 2-2, the number of cyclist injuries is expected to increase by 0.87% with a 1.0% increase in bicycle flows - regression coefficient of 0.869, credible interval (C.I.) [0.765 - 1.013]. However, holding other factors constant, the risk faced per cyclist decreases as the number of cyclists increases. In terms of motor-vehicle flows, both right turn (regression coefficient of 0.240, C.I. [0.153; 0.307]) and left turn (regression coefficient of 0.185, C.I. [0.106; 0.279]) movements have significant effects on cyclist injuries whereas through moving vehicles do not and therefore have been omitted from the final results. A 1.0% increase in right turn and left turn motor-vehicle flows are expected to cause a 0.24% and a 0.19% increase in cyclist injury occurrence, respectively.

The association between bicycle flows and the number of cyclist injuries highlights the importance of targeting intersections with high bicycle volumes. This work suggests that particular
attention be placed on right turns which should either be reduced or treated with specific countermeasures at intersections. Typical countermeasures include the reduction of turning radii, the implementation of an exclusive bicycle and pedestrian signal phase, bicycle boxes, etc. Restricting turning vehicular movements is a common practice in cities including Montreal, however it may simply move the problem to neighbouring intersections and can have negative impacts on network connectivity, travel times and delays. This countermeasure may however be justifiable at intersections with very high cyclist flows.

Other geometry and built environment attributes associated with injury occurrence are the presence of bus stops located at the intersection, crosswalk length and median presence. According to the elasticities, the presence of bus stops is associated with a 40% increase in cyclist injury occurrence - regression coefficient of 0.519, C.I. [0.196; 0.842]. In Montreal, bus stops are mostly located on major roads (arterials). Intersections with bus stops (70% of the sample) are busier, with more complex motor-vehicle and cyclist manoeuvring. Also, a 1.0% increase in the total crosswalk length at an intersection is associated with a 0.58% increase in injuries (regression coefficient of 0.009, C.I. [0.004; 0.014]). In other words, as the distance that cyclists need to cross increases so does the likelihood of them being involved in a crash. Retrofitting strategies such as curb extensions at intersections reduce road and therefore crossing lengths. The presence of a raised median at an intersection reduces injury occurrence by over 42% (regression coefficient of −0.351, C.I. [−0.640; −0.037]). Raised medians are found along at least one approach in 47% of the intersections in this study. Medians place constraints on motor-vehicle movements and can provide a refuge for cyclists who may have run out of time to safely cross the intersection. It is also important to mention that the presence of bicycle lanes or cycle tracks at intersections was also tested. However, they were not statistically significant with 95% credible intervals and included the zero value (C.I. [−0.12, 1.11] for bicycle lanes, and C.I. [−0.28, 0.85] for cycle tracks). Only a small number of intersections in this study have bicycle facilities, 29 with a bicycle lane and 27 with a cycle track, which may explain why they were not found to be significant. Therefore, there is not enough evidence to establish a positive (or negative) association between bicycle facility presence and injury frequency at signalized intersections.
2.6.2. Factors Associated with Bicycle Activity

A variety of built environment characteristics were tested to see their effects on bicycle activity. Employment, presence of schools, metro stations, land use mix, length of bicycle facilities and area of commercial spaces were found to have a significant and positive effect on bicycle activity. According to the elasticities estimated at the mean values, the presence of schools within 400 m of an intersection increases bicycle activity through that intersection by 25% (regression coefficient of 0.288, C.I. [0.258; 0.313]). The effect of metro stations on bicycle activity can be interpreted as a direct association, through mode transfers during a given trip (which can be marginal). However, an indirect association is more likely. Neighbourhoods with high transit use also tend to have more walking and cycling and, in Montreal, central neighbourhoods with metro stations have a higher bicycle mode share than more peripheral areas. Intersections with three approaches are expected to have fewer cyclists than intersections with four approaches (with an elasticity of 0.77). This factor can be seen as a proxy for intersection connectivity. Not surprisingly, the presence of bicycle facilities near an intersection has an important effect on bicycle activity. Intersections near bicycle facilities have a much higher concentration of bicycle flows with an elasticity of 0.288 and C.I. [0.131, 0.146]. Despite the outcome not being reported, a model was estimated using dummy variables to represent the presence of a bicycle lane or a cycle track at each intersection. The elasticities were found to be 0.37 and 0.23 for the presence of bicycle lanes and cycle tracks, respectively.

These results highlight the importance of indirect links between built environment and cyclist injuries. Changes in the built environment in the vicinity of an intersection can dramatically affect bicycle activity. For instance, the construction of a new school or bicycle facility will generate more bicycle traffic. This means that without the appropriate interventions implemented simultaneously with changes in the built environment, the number of cyclist injuries is expected to go up. One of the advantages of the joint analysis of cyclist injuries and bicycle activity is that the indirect impact of built environment changes on injuries can be estimated. To illustrate this, for example, a 10% increase in employment within 400 m of an intersection would result in an increase of 4.39% in bicycle flows which represents an increase of 3.8% (0.0439 × 0.869) in cyclist injuries at that intersection.
Finally, the covariance term between the injury and activity models was found to be significant. This demonstrates the importance of modeling these two outcomes simultaneously, taking into account the endogeneity problem – that unobserved factors affect both outcomes.

2.6.3. Cyclist Injury Risk along Corridors

The posterior expected number of injuries and the injury rates per million cyclists were obtained and computed from the model results for all intersections – Equations 2 and 3. Corridors were constructed from the intersections if a minimum of five intersections with data and intersections no more than 3 km apart were available. 51 corridors were constructed and injury rates and risk were computed by averaging these safety indicator values over all intersections belonging to the same corridor. The corridors were then ranked, from most to least dangerous, according to injury rates and injury frequency. These outcomes are reported in Figure 2-2 and Table 2-3. Figure 2-2 shows the corridors with the thicker lines representing the corridors with the greatest injury rates for cyclists. The rates are also shown through the use of colours ranging from green for safer to red for more dangerous intersections. Note that corridors far from downtown are those identified as having the greatest injury rates for cyclists (e.g., Boulevard Lacordaire and Boulevard Henri-Bourassa East in Saint Leonard) – one can also refer to the first part of Table 2-3. The worst corridors in terms of injury rates rank quite low in terms of observed and expected number of accidents. This result is not surprising given their relatively low bicycle volumes.

Table 2-3a lists the top 20 (out of 51) corridors in terms of risk (injury rates) while Table 2-3b shows the 20 worst corridors in terms of the posterior expected number of injuries. Overall Table 2-3 shows that corridors ranking high in terms of injury rates generally rank low in terms of injury frequency but rank even lower in terms of the number of cyclists riding along them. Only one corridor ranked in the top 20 in terms of injury rate (Table 2-3a) also ranked in the top 20 in terms of the expected number of injuries (Table 2-3b). On the other hand, corridors which are characterized as having a low risk for cyclists generally rank high in terms of accident numbers and rank very high in terms of cyclist numbers, therefore providing a low risk to each individual cyclist.

It is also worth mentioning that some of these corridors have either a bicycle path or cycle track along certain sections. Most importantly, Table 2-3a shows that the corridors with greater injury rates do not have any bicycle facility along them except for Rue Jean Talon Est, Avenue de
L’Église and Boulevard Thimens which each have a cycle track along a small section of the corridor. Table 2-3b shows that several corridors ranked in the top 20 in terms of injuries are partially served by a bicycle facility. Boulevard de Maisonneuve, which has a physically separated cycle track running along most of its length, ranks 49th in terms of risk, 15th and 9th in terms of observed and expected number of accidents respectively, while ranking 1st in terms of the number of cyclists. For corridors with cycle tracks (thirteen corridors), with bicycle paths (eight corridors) and with no bicycle facility, the average rank is 32, 40 and 21, respectively in terms of risk.

This corridor analysis revealed that corridors with high bicycle flows (exposure), mostly located in the central neighbourhoods, have lower individual risk of injury. This result may reflect the “safety in numbers” hypothesis or cyclist preferences – to ride in high numbers where safe designs have been implemented. However, corridors with high bicycle flows nevertheless have a greater likelihood of ranking high based on the absolute number of cyclist injuries. The presence of bicycle facilities translates into higher volumes and lower risk. Despite this, the injury frequency at intersections can still be reduced with appropriate designs such as adding exclusive green phases for cyclists at intersections with high turning movements, adding intersection refuges, eliminating bus stops in close proximity as well as reducing crossing length.
Figure 2-2 Injury rates along corridors
Table 2-3 Ranking of corridors

<table>
<thead>
<tr>
<th>Corridor</th>
<th>Average across corridors</th>
<th>Ranking of corridors</th>
<th>Bicycle Facility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Risk</td>
<td>Observed injuries**</td>
<td>Expected injuries**</td>
</tr>
<tr>
<td>Lacordaire</td>
<td>5.17</td>
<td>0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>de la Verendrye</td>
<td>3.63</td>
<td>0.13</td>
<td>0.09</td>
</tr>
<tr>
<td>Notre-Dame</td>
<td>3.11</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>Dickson</td>
<td>2.54</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Langelier</td>
<td>2.32</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Henri-Bourassa Est</td>
<td>2.25</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>Cavendish</td>
<td>2.17</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>Marcel-Laurin</td>
<td>2.17</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Vieu</td>
<td>2.00</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>Newman</td>
<td>1.89</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>Maurice-Duplessis</td>
<td>1.88</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>Jean-Talon Est</td>
<td>1.87</td>
<td>0.11</td>
<td>0.08</td>
</tr>
<tr>
<td>Cote-Vertu</td>
<td>1.71</td>
<td>0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>Dollard</td>
<td>1.66</td>
<td>0.13</td>
<td>0.10</td>
</tr>
<tr>
<td>Industriel</td>
<td>1.63</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>Henri-Bourassa Ouest</td>
<td>1.62</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>de l'Eglise</td>
<td>1.58</td>
<td>0.19</td>
<td>0.10</td>
</tr>
<tr>
<td>Papineau</td>
<td>1.53</td>
<td>0.29</td>
<td>0.21</td>
</tr>
<tr>
<td>Thimens</td>
<td>1.50</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>Cote-Saint-Luc</td>
<td>1.47</td>
<td>0.05</td>
<td>0.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Corridor</th>
<th>Average across corridors</th>
<th>Ranking of corridors</th>
<th>Bicycle Facility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Risk</td>
<td>Observed injuries**</td>
<td>Expected injuries**</td>
</tr>
<tr>
<td>Atwater</td>
<td>1.45</td>
<td>0.47</td>
<td>0.36</td>
</tr>
<tr>
<td>Jeanne-Mance</td>
<td>0.84</td>
<td>0.39</td>
<td>0.34</td>
</tr>
<tr>
<td>Viger</td>
<td>1.16</td>
<td>0.42</td>
<td>0.34</td>
</tr>
<tr>
<td>Saint-Laurent</td>
<td>0.99</td>
<td>0.47</td>
<td>0.32</td>
</tr>
<tr>
<td>Ontario</td>
<td>0.72</td>
<td>0.42</td>
<td>0.31</td>
</tr>
<tr>
<td>Saint-Denis</td>
<td>1.21</td>
<td>0.39</td>
<td>0.30</td>
</tr>
<tr>
<td>Rene-Levesque</td>
<td>1.09</td>
<td>0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>Parc</td>
<td>0.68</td>
<td>0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>de Maisonneuve</td>
<td>0.56</td>
<td>0.18</td>
<td>0.24</td>
</tr>
<tr>
<td>Amherst</td>
<td>0.91</td>
<td>0.28</td>
<td>0.24</td>
</tr>
<tr>
<td>Sherbrooke</td>
<td>0.75</td>
<td>0.21</td>
<td>0.23</td>
</tr>
<tr>
<td>Lorimier</td>
<td>1.16</td>
<td>0.20</td>
<td>0.22</td>
</tr>
<tr>
<td>University</td>
<td>1.38</td>
<td>0.15</td>
<td>0.21</td>
</tr>
<tr>
<td>Papineau</td>
<td>1.53</td>
<td>0.29</td>
<td>0.21</td>
</tr>
<tr>
<td>Hotel-de-Ville</td>
<td>0.53</td>
<td>0.10</td>
<td>0.19</td>
</tr>
<tr>
<td>Montagne</td>
<td>0.72</td>
<td>0.15</td>
<td>0.19</td>
</tr>
<tr>
<td>Pins</td>
<td>1.28</td>
<td>0.17</td>
<td>0.19</td>
</tr>
<tr>
<td>Sainte-Catherine</td>
<td>1.12</td>
<td>0.21</td>
<td>0.18</td>
</tr>
<tr>
<td>Saint-Antoine</td>
<td>0.90</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>Frontenac</td>
<td>0.76</td>
<td>0.07</td>
<td>0.14</td>
</tr>
</tbody>
</table>

* AADB stands for average annual daily bicycle flow
** computed based on the average of the intersections per year and averaged over the corridor

43
2.7. CONCLUSION AND FUTURE WORK

This study focused on almost 650 intersections on the island of Montreal providing a wide range of motor-vehicle traffic and bicycle exposure as well as injury counts. This paper proposed a simultaneous Bayesian modeling framework to investigate geometric design and built environmental factors associated with both cyclist injuries and volumes. The applicability of this framework for the identification of high-risk corridors was also illustrated.

Among other results, it was found that cyclist injury occurrence is sensitive to changes in both bicycle and motor-vehicle flows. This is in accordance with past studies. Cyclist volumes have a strong association with injury occurrence - a 1.0% increase in bicycle flows would result in a 0.87% increase in number of injuries. In terms of motor-vehicle flows, right turning vehicles were found to have the greatest effect whereas the effect of through moving motor-vehicles was found to be insignificant. Several geometric design and built environmental factors are associated with cyclist safety. For instance, cyclist injuries are expected to increase at intersections with a bus stop and with increasing crosswalk length. The presence of a raised median on the other hand, is expected to reduce crashes. The presence of bicycle facilities at intersections was not found to be statistically associated with injury frequency but has been found to increase cyclist volumes. Not surprising, intersections with bicycle facilities have a significantly higher number of cyclists. This means that, after controlling for other factors, intersections with bicycle facilities, with higher cyclist volumes, are expected to witness greater injury frequency but lower injury rates.

This study also demonstrated the important direct and indirect effects of built environment on cyclist activity and safety. Changes in the built environment are expected to cause direct changes in bicycle volumes and therefore indirect changes in injury frequency and injury risk at intersections. For instance, after the installation of a new bicycle facility crossing an intersection, bicycle flows are expected to grow as will the number of injuries without appropriate countermeasures. These results highlight the importance of improving safety standards when modifications are made to the geometry or built environment in the vicinity of an intersection. Also, as hypothesized, the correlation between the error terms due to unobserved factors was found to be significant therefore confirming the presence of endogeneity.

According to the descriptive corridor analysis, corridors ranking high in terms of injury rates generally rank low in terms of injury frequency. In other words, corridors with high bicycle
volumes, located mainly in the central neighbourhoods of Montreal, have a higher number of injuries but tend to have a lower risk per cyclist. The inverse association between injury risk and cyclist volume is commonly described as "safety in numbers". A similar phenomenon is well known in preventive medicine as the "preventive paradox" (Rose, 1992, 1985). For many diseases, high-risk individuals only provide a minority of all cases, and most cases come from the low- or average-risk population. In other words, there are more injured cyclists at low risk intersections and corridors because there are much more cyclists in these locations. Injury prevention strategies targeting locations at high risk can have a relatively small effect. However, the implementation of preventive intervention in low- or average-risk corridors may provide great public health benefits.

Although this study uses a large sample of intersections, this sample is not representative of the total population of intersections and therefore this presents a limitation of this work. As part of future work, a larger sample of intersections will be used to validate these results. This will help to draw more solid conclusions related to the role of bicycle facilities at signalized intersections. A random inventory of non-signalized intersections is currently in the process of being built and similar hypotheses will therefore be tested on a larger and more representative sample of all urban intersections in Montreal. Also, bicycle injury data provided by police reports will be used to validate these results. Before-after studies to evaluate the effectiveness of bicycle facilities (cycle tracks) are missing in the literature. More complex models will be attempted to investigate the potential presence of spatial autocorrelation. Finally, cyclist injuries occurring along road segments, between intersections, will be included in order to thoroughly study corridors in Montreal.

2.8. ADDENDUM

Some additional results that were not presented in this chapter are reported in Appendix A - Non-Signalized Intersections. These results come from replicating this chapter's analysis but for non-signalized intersections. Appendix A presents the cyclist injury and bicycle activity results in Table A-1. Appendix A also provides a discussion of these results.

2.9. ACKNOWLEDGMENTS

We acknowledge the financial support provided by the “Programme de recherche en sécurité routière” financed by FQRNT-MTQ-FRSQ. We would also like to thank the Montreal Department
of Public Health for providing injury data and the city of Montreal for providing motor-vehicle and cyclist traffic data. All remaining errors and the views expressed in this research are, however, solely ours. We would also like to thank the reviewers for providing great feedback helping us to improve this paper.

2.10. REFERENCES


**Link between Chapter 2 and Chapter 3**

Cyclists are not the only mode of transport using the roads and intersections in the urban road network of Montreal, but rather these facilities are shared by cyclists, pedestrians and motor-vehicles. Road safety is an important issue for all road users and it is for this reason that analysis needs to be shifted to a multimodal perspective. Accordingly, the following chapter focuses on the exposure, geometric design and built environment factors affecting the injury occurrence for three modes: cyclists, pedestrians and motor-vehicles, who are using the same facilities, which in this work are signalized and non-signalized intersections. It is important to incorporate all modes sharing the same space or infrastructure in order to identify whether or not a specific location is dangerous and for which modes. Typically, studies only focusing on one mode, for instance pedestrians, may identify the hazardous locations and implement safety treatments at the top-ranked sites. Without knowing the safety of cyclists or motor-vehicle occupants at these locations, the implemented solutions may have only solved part of the issue or worsened the location for the other modes. Based on the same intersection inventory for signalized intersections and a new inventory for non-signalized intersections, the next chapter studies the safety of all three modes at both signalized and non-signalized intersections. This chapter builds on the previous chapter by adapting the same Bayesian modeling framework to simultaneously model injury and activity outcomes, to account for the potential issues of endogeneity, for signalized and non-signalized intersections from a multimodal perspective.
Chapter 3

Multimodal Injury Risk Analysis between Road Users at Signalized and Non-Signalized Intersections
Chapter 3: Multimodal Injury Risk Analysis between Road Users at Signalized and Non-Signalized Intersections

Jillian Strauss\textsuperscript{a}, Luis Miranda-Moreno\textsuperscript{a} and Patrick Morency\textsuperscript{b}
\textsuperscript{a}Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, Canada
\textsuperscript{b}Montreal Department of Public Health, Montreal Health and Social Service, Montreal, Canada

This work was presented at the Transportation Research Board 93\textsuperscript{rd} Annual Meeting in Washington, D.C., 2014 and at the Canadian Multidisciplinary Road Safety Conference in Vancouver, British Columbia, 2014 where it was awarded the first place paper by the Insurance Bureau of Canada (IBC). This paper has also been published in Accident Analysis and Prevention.

3.1. ABSTRACT

This paper proposes a multimodal approach to study safety at intersections by simultaneously analyzing the safety and flow outcomes for both motorized and non-motorized traffic. This study uses an extensive inventory of signalized and non-signalized intersections on the island of Montreal, Quebec, Canada, containing disaggregate motor-vehicle, cyclist and pedestrian flows, injury data, geometric design, traffic control and built environment characteristics in the vicinity of each intersection. Bayesian multivariate Poisson models are used to analyze the injury and traffic flow outcomes and to develop safety performance functions for each mode at both facility types. After model calibration, contributing injury frequency factors are identified. Injury frequency and injury risk measures are then generated to carry out a comparative study to identify which mode is at greatest risk at intersections in Montreal. Among other results, this study identified the significant effect that motor-vehicle traffic imposes on cyclist and pedestrian injury occurrence. Motor-vehicle traffic is the main risk determinant for all injury and intersection types. This highlights the need for safety improvements for cyclists and pedestrians who are, on average, at 14 and 12 times greater risk than motorists, respectively, at signalized intersections. Aside from
exposure measures, this work also identified some geometric design and built environment characteristics affecting injury occurrence for cyclists, pedestrians and motor-vehicle occupants.

3.2. INTRODUCTION

Urban mobility and safety for all modes of transportation are key elements in the development of sustainable cities. Reaching this goal requires the implementation of multimodal approaches and a shift towards non-motorized modes of transportation, namely walking and cycling.

Despite the health and environmental benefits, choosing walking or cycling as the desired mode exposes cyclists and pedestrians to safety risks. Motor-vehicle collisions are an important cause of death and injury worldwide. Over 2000 people, motor-vehicle occupants, motorcyclists, cyclists and pedestrians are killed each year on Canadian roads (Transport Canada, 2010). The likelihood of fatal and severe injuries are, in general, greater for pedestrians and cyclists than for motor-vehicle occupants (Beck et al., 2007). In 2010, 73% of road injuries occurred in urban areas in Canada (Transport Canada, 2010). In urban settings, the majority of pedestrians and cyclists are injured at intersections. Urban intersections are a complex area of the road network where many different interactions can occur between motor-vehicles, cyclists and pedestrians. Road safety at urban intersections, for all modes, is therefore an important issue.

Despite the important body of literature, important gaps still exist in data collection methods, data integration and modeling approaches. In particular multimodal safety approaches at the city level are missing in the literature. Previous research has described the overall number of road injuries or fatalities, without making any distinction between the road user types. Additionally, past studies have considered single outcomes, such as the safety of cyclists, pedestrians, or motorized traffic and few have integrated all these modes. Also, the effect of geometric design and built environment characteristics related to the layout and location of the intersections being studied, have not been tested.

Motor-vehicle occupant and pedestrian injury risk at intersections have been studied for decades, but empirical evidence for cyclist risk at intersections is scarce. Furthermore, little is known about the contributing environmental factors associated with injury risk across users (cyclist, pedestrian and motor-vehicle). Instead of focusing only on one mode of travel and identifying sites in need of improvement, this study considers cyclists, pedestrians and motor-vehicle occupants and applies a multimodal approach. A multimodal approach can help build more
comprehensive safety portraits, since the injury risk of each travel mode is simultaneously determined. Thus, if intersection characteristics are found to increase crashes for cyclists and pedestrians, for example, safety improvements can be carried out for both modes at once. A multimodal approach may ease the task of selecting sites for safety improvement interventions as well as potentially provide a more economically viable solution, compared to separated analysis and interventions for pedestrians and cyclists. Finally, empirical evidence regarding the risk imposed by motor-vehicles on non-motorized users at intersections is mostly limited to signalized intersections, which only represent a minority of urban intersections. The inclusion of non-signalized intersections would help to better understand the safety issues of non-motorized users in urban settings and provide a complete portrait of all intersections within a city.

In this paper, we seek to provide additional evidence on the abovementioned research issues using as an application environment the city of Montreal, Quebec, Canada. This paper aims to: i) develop injury occurrence models, ii) estimate the risk for all road users using a Bayesian modeling approach, iii) carry out a comparative analysis between the injuries, levels of flow and risk across the three modes (cyclist, pedestrian and motor-vehicle) and two facility types (signalized and non-signalized intersections), and iv) investigate the impact of motor-vehicle traffic on cyclist and pedestrian safety at intersections.

This paper is broken down into several sections. Section 2 offers a literature review. Section 3 describes the methodology followed by the site selection and data in section 4. Section 5 discusses the results and section 6 presents the conclusions drawn from this study as well as directions for future work and limitations.

3.3. LITERATURE REVIEW

Road safety in urban environments is a topic that has been growing in the literature. In recent years, this extensive literature has dealt with many issues ranging from data collection methods, data integration and modeling approaches. Despite all these efforts, multimodal safety approaches at the city level are missing in the literature. Some studies have looked at multiple injury severity outcomes of motor-vehicles (minor, major and fatalities, for instance). However, to our knowledge, no studies have looked at multiple road user outcomes (cyclist, pedestrian and motorized outcomes). Most studies have looked at single crash outcomes.
For instance, cyclist safety studies examining the factors affecting crash occurrence have been published (Brüde and Larsson, 1993; Elvik, 2009; Jacobsen, 2003; Miranda-Moreno et al., 2011; Wang and Nihan, 2004). To date, most cyclist safety studies have been carried out in European and Asian cities and while a few have been carried out in the United States and in Australia, these have mainly focused on cyclist injuries at the city or town level and did not focus on intersections (junctions) as the unit of study (Elvik, 2009). Pedestrian and motor-vehicle safety studies in North America are not as rare. For studies focusing on pedestrian-vehicle collisions at intersections, we can refer to the works of Cameron (1982), Brüde and Larsson (1993), Lyon and Persaud (2002), Shankar et al. (2003), Lee and Abdel-Aty (2005), Harwood et al. (2008) and Miranda-Moreno et al. (2011) and for vehicle-vehicle collisions, refer to Boufous et al. (2008) and Chin and Quddus (2003).

Among the studies focusing on intersections as the unit of analysis, the majority of these considered signalized and not non-signalized intersections. Among the studies addressing safety at non-signalized intersections in urban environments, we can mention the work of Sayed and Rodriguez (1999); few studies however have provided safety portraits and models for both types of intersections in the same city.

While these previous studies are useful for identifying the factors contributing to cyclist, pedestrian and motor-vehicle injury occurrence, these have been modeled separately and few attempts have been made to combine these into a multimodal approach. Some studies in the literature have addressed level-of-service, delay or comfort for cyclists, pedestrians and motor-vehicles from a multimodal perspective (Dowling et al., 2008; Guttenplan et al., 2001) however safety has received less attention in the literature.

One study in the United States compared the injury risk between different modes (Beck et al., 2007) based on the observed injuries and exposure in terms of daily person-trips. This study did not consider factors other than exposure which may also be determinants of injury occurrence. The results identified that motorcyclists, cyclists and pedestrians are more at risk of being fatally injured (58.3, 2.3 and 1.5 times more, respectively) compared to motor-vehicles. However in terms of injuries, motor-vehicle occupants account for the majority of fatal and non-fatal injuries. Another study carried out in the United Kingdom sought to investigate the risk of cycling in absolute terms and with respect to other modes of transportation (Wardlaw, 2002). Per kilometre travelled, pedestrians witnessed a greater level of fatalities due to motor-vehicle collisions than
cyclists. Wardlaw (2002) emphasizes that cycling is not a dangerous mode of transportation and the belief that it is in fact dangerous is not based on evidence.

In addition to modeling injury occurrence, it is important to obtain a basis for which the safety at each location and for each mode can be compared. Recent studies have defined and applied a Bayesian Hierarchical approach to identify hazardous locations (Brijs et al., 2007; Miaou and Song, 2005; Schlüter, 1997; Tunaru, 2002; Van den Bossche et al., 2002). Many applications of the multivariate Poisson hierarchical model with covariates and time-space random effects have also been documented in the road safety literature. For a comprehensive literature review, refer to Lord and Mannering (2010). Bayesian models are very popular in the road safety literature since posterior risk estimates can be computed easily. Risk estimates can then be used to identify intersections, corridors or areas at high risk.

Among the most popular posterior risk estimates are the posterior mean number of injuries, the posterior injury rate, the potential for injury reduction, the posterior distribution of ranks and the posterior probability that a site is the worst (Miaou and Song, 2005; Miranda-Moreno, 2006; Schlüter, 1997). Despite this extensive literature, little effort has been made to use Bayesian methods to identify dangerous intersections for cyclists and pedestrians while also correcting for explanatory variables such as traffic conditions and controls as well as geometric design and built environment characteristics. In addition to identifying hotspots for specific modes or for total road accidents, it is important to know which modes are most at risk and where. If, for example, locations where cyclists are most at risk are locations where pedestrians are also at risk, this influences what type of treatments should be proposed and implemented. The task becomes more complex when the high-risk sites for different modes do not overlap.

In response to the research questions and gaps in the literature, this study applies a multimodal approach to study cyclist, pedestrian and motor-vehicle safety at both signalized and non-signalized intersections using Bayesian multivariate Poisson models.

3.4. METHODOLOGY

Based on our previous work (Strauss et al., 2013a), a bivariate Bayesian Poisson model is used to represent injury outcome and flow simultaneously since unobserved factors may be affecting both outcomes and potentially generating an endogeneity problem. This model is formulated as:
\[ Y_{ik} \mid \theta_{ikY} \sim \text{Poisson}(t_{ikY}) \]
\[ Z_{ik} \mid \theta_{ikZ} \sim \text{Poisson}(\theta_{ikZ}) \]

With
\[ \theta_{ikY} = V_i^{\alpha_1}Z_{ik}^{\alpha_2}\exp(\beta_0 + \beta_1x_{ik1} + \ldots + \beta_jx_{ikj} + \varepsilon_{ikY}) \]
\[ \theta_{ikZ} = \exp(\gamma_0 + \gamma_1w_{ik1} + \ldots + \gamma_jw_{ikj} + \varepsilon_{ikZ}) \]
\[ \varepsilon_i \mid T \sim N_m(0, T), \]

Where, \( Y_{ik} \) is the number of injuries observed at a given intersection \( i \) (\( i=1,..n \) intersections) for a given mode \( k \) (cyclist, pedestrian or motor-vehicle) during a period of time \( t_i \). \( Z_{ik} \) is the average annual daily flow at the same intersection for mode \( k \) (cyclist or pedestrian) calculated by applying expansion factors to the manual counts. \( Y_{ik} \) and \( Z_{ik} \) are Poisson distributed with \( \theta_{ikY} \) and \( \theta_{ikZ} \) denoting the mean injury frequency and the mean daily flow of mode \( k \), respectively. \( V_i \) represents traffic flows, either aggregated to the entire intersection or disaggregated into left turn, right turn and through movements at each intersection. Injuries are modeled as a function of flows, \( V_i \) and \( Z_{ik} \), as well as some geometric design and built environment characteristics, represented by \( x \) (such as the presence of bus stops and raised medians, number of lanes, crosswalk length, commercial entrances and exits, all-red and half-red pedestrian and cyclist phases as well as others). Cyclist and pedestrian flow (\( Z_{ik} \)) are modeled as a function of mostly built environment characteristics but also some geometric design attributes, \( w \) (such as employment, presence of schools and metro stations, land use mix and area of commercial land use, length of bicycle facilities, presence of three approaches (versus four) as well as others). Both models have an error term, \( \varepsilon_1 \) and \( \varepsilon_2 \), which are correlated. \( N_m() \) stands for the bivariate Normal distribution with mean vector 0 and covariance matrix \( T \). \( \alpha_1, \alpha_2, \beta=(\beta_0,\ldots, \beta_j) \) and \( \gamma=(\gamma_0,\ldots, \gamma_j) \) are model regression parameters to be estimated from the data. Priors on these parameters are assumed to follow non-informative or informative Gaussian (Normal) distributions. Non-informative priors are assumed to have a mean equal to 0 and a very large variance, e.g. \( \beta_j \sim N(0,1000) \). Informative priors can be assumed for certain parameters with prior knowledge. Also, the regression parameters are mutually independent. Note that this model can easily be extended to a multivariate approach (e.g., five simultaneous outcomes: cyclist and pedestrian injuries and flows as well as motor-vehicle occupant injuries).

OpenBUGS (Bayesian inference using Gibbs sampling), an open source software, uses Gibbs sampling for posterior inference after model formulation. The parameter estimates (posterior means, standard deviations, t-statistics and credible intervals) and the DIC (Deviance Information
Criterion) value are sampled after burning the first 150,000 updates then sampling the next, 10,000. After proposing the simultaneous model, these results are applied for cyclist, pedestrian and motor-vehicle injury occurrence at both signalized and non-signalized intersections to compare the safety between modes and intersection types. Two criteria are used for the comparison: 1) posterior mean of injury frequency at intersection \( i \), \( \theta_{iy} \), and 2) injury rates. The first criterion, \( \overline{\theta}_{iy} \), is perhaps the most popular criterion in the safety literature. It is generally defined as shown in Equation 1:

\[
\overline{\theta}_{iy} = E[\theta_{iy} | y_i] = \int_{\theta_{iy}} \theta_{iy} p(\theta_{iy} | y_i) d\theta_{iy}
\] (1)

The second criterion is injury rates, also referred to as risk. Based on the posterior mean of injury frequency at intersection \( i \), the posterior injury rate per million cyclists, pedestrians and motor-vehicles per unit of time can be computed. Risk is denoted by \( R_{iy} \) and defined as shown in Equation 2:

\[
R_{iy} = \overline{\theta}_{iy} \times 10^6 / 365 t_i Z_i
\] (2)

### 3.5. SITE SELECTION AND DATA

#### 3.5.1. Site Selection

The island of Montreal, Quebec, Canada is used as the application environment. Montreal has a population of just under 1.9 million people (2011 census) and is characterized by an important modal share of transit and active modes of transportation. Transit trips represent 22%, while walking and bicycle trips represent 10% (the remaining 68% represent the car share). This study focuses on 647 signalized and 435 non-signalized intersections for which both outcomes are known: cyclist, pedestrian and motor-vehicle injuries and flows. It is important to mention that the signalized intersections in this sample include all combinations of arterial, collector and local streets intersecting (17% arterial-arterial, 22% arterial-collector, 41% arterial-local, 3% collector-collector, 11% collector-local and 6% local-local). As a starting point, the non-signalized intersections in this sample all include arterial, collector or local streets intersecting with local streets. In other words, all non-signalized intersections in this study are classified as arterial-local (11%), collector-local (21%) or local-local (68%), selected by a random representative sample. Figure 3-1 shows a map of the island of Montreal and the samples of signalized (represented by green circles) and non-signalized intersections (represented by red squares) studied. These
intersections were selected since an extensive inventory was constructed at McGill University which combines a wide variety of data including manual counts for cyclists, pedestrians and motor-vehicles, injury data, geometric design, traffic control and built environment data.

3.5.2. Signalized and Non-Signalized Intersection Inventories

a. Manual Count Data

In 2008 and mostly in 2009, the city of Montreal carried out eight-hours (6:00 a.m.–9:00 a.m., 11:00 a.m.–1:00 p.m. and 3:30 p.m.–6:30 p.m.) of manual bicycle, pedestrian and motor-vehicle counts at signalized intersections. These flows were collected on weekdays during the cycling season, from April 1st and November 30th.

Summer and fall of 2012 (May to November), we set out on a similar endeavor to collect traffic counts at non-signalized intersections. Due to limited time and resources to carry out counts at over 400 intersections, one full hour of counts during either morning (6:00 a.m.–10:00 a.m. or until 11:00 a.m. in the central business district) or evening peak (3:00 p.m.–7:00 p.m.) per intersection was obtained and all counts were carried out on a weekday during the cycling season.

Using expansion factors developed specifically for Montreal, bicycle and pedestrian flows could be adjusted for the hour(s), day and month when the counts were taken to obtain average
annual daily bicycle and pedestrian volumes passing through each intersection. These factors have been reported in Strauss (2012b). The traffic flows were also expanded using the expansion factors for motor-vehicles provided by the city of Montreal. In regression analysis, the average annual daily bicycle, pedestrian and motor-vehicle flows were transformed using the natural logarithm (Ln).

Summary statistics of this flow data is provided in Table 3-1. This table highlights the differences in the magnitude of bicycle, pedestrian and motor-vehicle flows at signalized and non-signalized intersections. The mean number of average annual daily cyclists (445) and pedestrians (1578) is 2.4 and 2.2 times greater at signalized than at non-signalized intersections, respectively (186 cyclists and 707 pedestrians). For motor-vehicles, the mean average annual daily flow at signalized intersections is 5.4 times greater than at non-signalized intersections (24624 and 4578 for signalized and non-signalized).

Table 3-1 Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Signalized Intersections</th>
<th>Non-Signalized Intersections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyclist injury count 2003-2008</td>
<td>0.63 1.32 0 20</td>
<td>0.13 0.47 4 0</td>
</tr>
<tr>
<td>Pedestrian injury count 2003-2008</td>
<td>1.15 1.88 0 16</td>
<td>0.067 0.315 3 0</td>
</tr>
<tr>
<td>Motor-vehicle injury count 2003-2008</td>
<td>4.56 6.37 0 58</td>
<td>0.287 0.881 7 0</td>
</tr>
<tr>
<td>Average annual daily bicycle flows</td>
<td>445 714.9 1.66 6433</td>
<td>186 400.0 0 2955</td>
</tr>
<tr>
<td>Average annual daily pedestrian flows</td>
<td>1578 3531.9 0 40958</td>
<td>707 1387.4 0 13505</td>
</tr>
<tr>
<td>Average annual daily motor-vehicle flows</td>
<td>24624 12550.1 3751 84386</td>
<td>4578 5659.4 13 31745</td>
</tr>
<tr>
<td>Average raw risk for cyclists</td>
<td>2.21 13.97 0.00 304.41</td>
<td>0.50 4.12 0.00 9.51</td>
</tr>
<tr>
<td>Average raw risk for pedestrians</td>
<td>2.81 16.50 0.00 228.31</td>
<td>0.07 0.60 0.00 8.62</td>
</tr>
<tr>
<td>Average raw risk for motor-vehicle occupants</td>
<td>0.09 0.12 0.00 1.30</td>
<td>0.03 0.14 0.00 2.48</td>
</tr>
<tr>
<td>Presence of bus stops</td>
<td>0.71 0.46 0 1</td>
<td>0.31 0.46 0 1</td>
</tr>
<tr>
<td>Presence of raised median</td>
<td>0.47 0.5 0 1</td>
<td>0.09 0.28 0 1</td>
</tr>
<tr>
<td>Total number of lanes (sum for all approaches)</td>
<td>6.9 2.6 0 16</td>
<td>3.3 1.2 0 12</td>
</tr>
<tr>
<td>Total crosswalk length (metres, sum for all approaches)</td>
<td>68.5 24.1 9.29 245</td>
<td>Not currently available</td>
</tr>
<tr>
<td>400 metres employment ('000)*</td>
<td>1.5 0.97 0.06 4.7</td>
<td>3.3 7.8 0.016 71.24</td>
</tr>
<tr>
<td>400 metres presence of schools*</td>
<td>0.56 0.5 0 1</td>
<td>0.51 0.5 0 1</td>
</tr>
<tr>
<td>800 metres metro (subway) stations*</td>
<td>1.5 1.8 0 7</td>
<td>0.6 1.4 0 7</td>
</tr>
<tr>
<td>800 metres land use mix*</td>
<td>0.67 0.15 0 0.92</td>
<td>0.55 0.19 0 0.93</td>
</tr>
<tr>
<td>50 metres area of commercial land use (square metres '000)*</td>
<td>0.89 1.4 0 7.8</td>
<td>0.25 0.74 0 5.25</td>
</tr>
<tr>
<td>800 metres length of bicycle facilities (kilometres)*</td>
<td>2.1 1.7 0 7.9</td>
<td>2.29 1.66 0 9.50</td>
</tr>
<tr>
<td>Presence of three approaches</td>
<td>0.21 0.41 0 1</td>
<td>0.78 0.42 0 1</td>
</tr>
<tr>
<td>Number of observations</td>
<td>647</td>
<td>435</td>
</tr>
</tbody>
</table>

* These variables were measured within either a 50, 150, 400 or 800 metre radius around the intersections
b. Injury Data

This study focuses on cyclist, pedestrian and motor-vehicle occupant (driver and passenger) injuries which occurred at intersections over a six-year study period from 2003 to 2008 obtained from ambulance services. Accidents are considered as having occurred at an intersection if they are within 15 metres of the centre point of the intersection. Although ambulance data may be biased towards more severe injuries, in Montreal, this source of data identified more cyclist injuries than police reports. From 2003 to 2007, the ambulance data reported on average 1,011 injuries per year whereas the police report data only reported 787 injuries per year. However for pedestrians and motor-vehicles, ambulance data has fewer records than the police report data. Injury data is provided at the level of the individual and not at the level of the crash. A crash involving two cyclists (or pedestrians) would be considered as two separate injuries, however this situation arises in a very small percentage of cases in this dataset.

The island of Montreal has a total of 18,116 intersections, 2,074 signalized and 16,042 non-signalized. Figure 3-2 shows the distribution of the average number of cyclist and pedestrian injuries across different ranges of motor-vehicle occupant injuries for the entire population of signalized and non-signalized intersections in Montreal from 2003-2008. For both types of intersections, there is a clear association between the number of injuries for all modes. In other words, intersections with a high number of motor-vehicle occupant injuries, also witnessed high numbers of both cyclist and pedestrian injuries. This implies that intersections which witnessed a high number of injuries, witnessed a high number of cyclist, pedestrian and motor-vehicle injuries.

![Graph showing distribution of injuries at signalized and non-signalized intersections]

**Figure 3-2 Distribution of Injuries at Signalized (left) and Non-Signalized (right) Intersections**
These graphs also show the same pattern for signalized and non-signalized intersections, however the number of injuries at signalized intersections is much higher for all three modes.

Based on this data and the intersections studied, the average number of cyclist injuries which occurred at signalized intersections (0.63) is 4.8 times greater than at non-signalized intersections (0.13). For pedestrians, the mean injury occurrence is about 17.2 times greater at signalized (1.15) than at non-signalized intersections (0.067) and for motor-vehicles, 15.9 times more injuries occurred at signalized intersections (4.56) than at non-signalized intersections (0.287). There was at least one pedestrian injured at 45% of signalized intersections but only at 5% of non-signalized intersections. Similarly for cyclists, there was at least one cyclist injured at 36% of signalized intersections but only at 9% at non-signalized intersections. Also, the highest observed number of pedestrian and cyclist injuries at non-signalized intersections is 3 and 4, respectively, and some signalized intersections witness as high as 20 and 16 cyclist and pedestrian injuries, respectively. Table 3-1 also shows that at both signalized and non-signalized intersections, motor-vehicles have by far the greatest flows compared to cyclists and pedestrians. The individual raw risk (crash rates), computed using Equation 2 with observed injuries, is by far the lowest for motor-vehicles. In other words, non-motorized traffic, on average, is at higher risk of injury than motorized users at both signalized and non-signalized intersections despite that motorists represent 72% of all crashes at signalized and 60% of all crashes at non-signalized intersections.

c. Intersection Geometric Design and Traffic Control Data

An important data collection campaign was undertaken by students at McGill University during the summer and fall seasons of 2010 to 2012. Our team collected a rich inventory of signalized and non-signalized intersections in Montreal. The data collected includes a wide variety of geometric design and traffic control characteristics such as: number of approaches and lanes, type of traffic signals, pedestrian/bicycle phasing, left turn lanes and phasing, presence and width of medians, presence and type of bicycle facilities, crosswalk length and so on. Students visited each intersection in pairs with specific data collection sheets and an odometer. Summary statistics of this data is also provided in Table 3-1.
d. Built Environment Characteristics

Built environment characteristics such as land use, urban form and bicycle facility characteristics have been provided by different sources: Statistics Canada, DMTI Spatial Inc., Société de transport de Montréal (STM) and Agence métropolitaine de transport (AMT). This data includes population, employment, income, land use, presence of metro (subway) stations, bus stops, street typology, other demographics and road and transit characteristics. These variables were extracted for four different buffer dimensions: 50 metres, 150 metres, 400 metres and 800 metres to evaluate the impact of these variables at different distances from the intersection. Although bicycle activity may be better predicted using larger buffer sizes, caution must be taken when selecting variables for the model since the proportion of correlated variables is very likely to increase with increasing buffer sizes which we witnessed with the current buffer dimensions.

3.6. RESULTS

3.6.1. Cyclist, Pedestrian and Motor-Vehicle Injury Models

The best regression outcomes, selected based on the best DIC (Deviance Information Criterion) values and significance of all variables (95% level), are presented in Table 3-2. This table shows the parameters chosen in the final injury models for each mode, including the value of the coefficient, the t-statistic (t-stat) and the credible interval. Refer to Strauss et al. (2013) for cyclist injury elasticities and refer to Strauss et al. (2012) and Miranda-Moreno and Fernandes (2011) for bicycle and pedestrian activity model results and elasticities, respectively.

The cyclist and pedestrian injury models identify the importance of cyclist, pedestrian and motor-vehicle flows on injury occurrence. At signalized intersections, cyclist injuries are expected to increase by more than 8% with a 10% increase in cyclist flows (Table 3-2a). Pedestrian injuries are expected to increase by more than 5.5% with a 10% increase in pedestrian flows (Table 3-2c). At non-signalized intersections, cyclist and pedestrian flows have a similar elasticity around 7% (Table 3-2b and d). In accordance with our hypothesis, motor-vehicle traffic increases injury occurrence for cyclists and pedestrians at signalized and non-signalized intersections. For cyclists, injuries are expected to increase by 2.4% and 1.85% with 10% increases in right turn and left turn motor-vehicle flows, respectively, at signalized intersections and increase by 2.6% with a 10% increase in total motor-vehicle flows through non-signalized intersections. For pedestrians, injuries
are expected to increase by 5.6% with a 10% increase in motor-vehicle flows at signalized intersections and by 4.16% with a 10% increase in motor-vehicle flows at non-signalized intersections.

The model results for motor-vehicles, Table 3-2e and f, further emphasize the positive association between motor-vehicle flows and injuries for all modes at intersections, signalized or not. For motor-vehicle occupants and cyclists, disaggregate motor-vehicle movements were significantly associated with injury occurrence.

In addition to exposure measures, some intersection attributes were found to have significant effects on injury occurrence. At signalized intersections, the presence of bus stops located at the intersection is associated with a 40% increase in cyclist injury occurrence and a 48% increase in motor-vehicle collisions. In Montreal, bus stops are most frequently located on major roads (arterials). Intersections with bus stops are busier, with more complex motor-vehicle and cyclist manoeuvring. A 10% increase in the total crosswalk length at an intersection would cause a 6% increase in cyclist injuries. In other words, as the distance that cyclists need to cross increases, so does the likelihood of them being involved in a crash. Retrofitting strategies such as curb and sidewalk extensions at intersections reduce road and therefore crossing lengths. The presence of a raised median at an intersection reduces cyclist injury occurrence by over 40%. Raised medians are found along at least one approach in 47% of signalized intersections in this study. Medians place constraints on motor-vehicle movements and can provide a refuge for cyclists who may have run out of time to safely cross the intersection. Intersections in close proximity to commercial entrances and exits are more dangerous for pedestrians than intersections without. Also the all-red (referring to an exclusive all-red phase in which the signal is red for all traffic from every approach and pedestrians can cross to and from any corner) and half-red (referring to a phase for which only a portion of it (for example, 5 seconds) provides an exclusive phase for pedestrians). This is done either by using extended red lights for cars or a green arrow with a pedestrian signal at the same time. Both the extended red light and green arrow prevent cars from turning and conflicting with crossing pedestrians) signal phases reduce pedestrian injuries at signalized intersections. The model results at signalized intersections also identify that fewer motor-vehicle collisions are expected at intersections with three entering approaches as opposed to intersections with four approaches due to a lower likelihood of conflicts.
Of the intersection attributes tested, only the total number of lanes entering the intersection was found to have a significant effect on cyclist injury occurrence at non-signalized intersections.

It is worth mentioning that correlation does exist between traffic flow and crosswalk length at intersections. To investigate the effect that these correlations have on the injury occurrence results, models were fitted to the data with and without crosswalk length for cyclists. The parameter coefficients were not found to be sensitive to the correlation between traffic flow and crosswalk length and the model with crosswalk length was found to have a better fit. These results show that the general conclusions are consistent, however, correlation should be taken into account when interpreting the outcomes.

Table 3-2 Cyclist, Pedestrian and Motor-Vehicle Injury Model Results for Signalized and Non-Signalized Intersections

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln bicycle flows</td>
<td>0.869</td>
<td>12.27</td>
<td>0.765</td>
<td>0.103</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln mv* right turn flows</td>
<td>0.240</td>
<td>6.11</td>
<td>0.153</td>
<td>0.307</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln mv* left turn flows</td>
<td>0.185</td>
<td>3.70</td>
<td>0.106</td>
<td>0.279</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence of bus stops</td>
<td>0.519</td>
<td>3.16</td>
<td>0.196</td>
<td>0.842</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total crosswalk length</td>
<td>0.009</td>
<td>3.56</td>
<td>0.004</td>
<td>0.014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raised median</td>
<td>-0.351</td>
<td>-2.29</td>
<td>-0.640</td>
<td>-0.037</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-10.08</td>
<td>-30.66</td>
<td>-10.66</td>
<td>-9.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>7134</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln bicycle flows</td>
<td>0.748</td>
<td>5.03</td>
<td>0.447</td>
<td>0.998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln mv* flows</td>
<td>0.261</td>
<td>2.66</td>
<td>0.057</td>
<td>0.415</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of lanes</td>
<td>0.192</td>
<td>1.81</td>
<td>-0.020</td>
<td>0.386</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-17.52</td>
<td>-38.33</td>
<td>-18.48</td>
<td>-16.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>3462</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln pedestrian flows</td>
<td>0.565</td>
<td>25.90</td>
<td>0.515</td>
<td>0.603</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln mv* flows</td>
<td>0.561</td>
<td>12.43</td>
<td>0.503</td>
<td>0.628</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All-red phase</td>
<td>-0.619</td>
<td>-2.18</td>
<td>-1.181</td>
<td>-0.092</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Half-red phase</td>
<td>-0.400</td>
<td>-2.51</td>
<td>-0.719</td>
<td>-0.085</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial entrances/exits</td>
<td>0.095</td>
<td>2.40</td>
<td>0.017</td>
<td>0.173</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.052</td>
<td>5.17</td>
<td>0.684</td>
<td>1.294</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>7754</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln pedestrian flows</td>
<td>0.702</td>
<td>3.440</td>
<td>0.286</td>
<td>1.094</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln mv* flows</td>
<td>0.416</td>
<td>2.040</td>
<td>0.039</td>
<td>0.783</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>-0.619</td>
<td>-2.18</td>
<td>-1.181</td>
<td>-0.092</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Half-red phase</td>
<td>-0.400</td>
<td>-2.51</td>
<td>-0.719</td>
<td>-0.085</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial entrances/exits</td>
<td>0.095</td>
<td>2.40</td>
<td>0.017</td>
<td>0.173</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.052</td>
<td>5.17</td>
<td>0.684</td>
<td>1.294</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>7754</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln mv* right turn flows</td>
<td>0.174</td>
<td>9.29</td>
<td>0.148</td>
<td>0.212</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln mv* left turn flows</td>
<td>0.163</td>
<td>8.01</td>
<td>0.121</td>
<td>0.195</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln mv* through flows</td>
<td>0.263</td>
<td>24.37</td>
<td>0.239</td>
<td>0.282</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence of bus stops</td>
<td>0.661</td>
<td>5.05</td>
<td>0.409</td>
<td>0.897</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three approaches</td>
<td>-0.350</td>
<td>-2.58</td>
<td>-0.626</td>
<td>-0.081</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-4.63</td>
<td>-56.31</td>
<td>-4.74</td>
<td>-4.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>2684</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
<th>Variable</th>
<th>Coefficient</th>
<th>T-stat</th>
<th>Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln mv* right turn flows</td>
<td>0.166</td>
<td>1.27</td>
<td>-0.103</td>
<td>0.399</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln mv* left turn flows</td>
<td>0.129</td>
<td>1.75</td>
<td>-0.030</td>
<td>0.247</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln mv* through flows</td>
<td>0.951</td>
<td>10.04</td>
<td>0.7868</td>
<td>1.103</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-11.70</td>
<td>-13.79</td>
<td>-13.00</td>
<td>-9.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>412</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* mv = motor-vehicle
The graphs in Figure 3-3 show the distribution of injuries at signalized (left) and non-signalized (right) intersections. The distribution of the observed (all 18,116 intersections) and expected (obtained from the models for the sample of 1,082 intersections (647 signalized and 435 non-signalized)) number of injuries by mode are similar. For both facility types, there are many more motor-vehicle injuries than cyclist and pedestrian injuries combined. Also, there are more pedestrian than cyclist injuries at signalized intersections and the reverse is true for non-signalized intersections.

![Figure 3-3 Percentage of Total Observed and Expected (in brackets) Injuries by Mode at Signalized (left) and Non-Signalized Intersections (right)](image)

**3.6.2. Expected Risk: Sample of Intersections**

Risk computations require cyclist, pedestrian and motor-vehicle exposure and therefore this analysis can only be carried out for the sample of intersections for which the flows for all three modes are available. Expected risk, as defined in Equation 2, is the expected number of injuries (obtained from the models) per million cyclists, pedestrians or motor-vehicle occupants per year. Summary statistics of expected risk are provided in Table 3-3 and the graphs in Figure 3-4 show the distribution of intersections according to risk, for the three modes of interest.

**Table 3-3 Summary Statistics of Expected Risk**

<table>
<thead>
<tr>
<th>Intersection</th>
<th>Expected Risk*</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signalized</td>
<td>Cyclist</td>
<td>1.186</td>
<td>0.981</td>
<td>0.015</td>
<td>5.826</td>
</tr>
<tr>
<td></td>
<td>Pedestrian</td>
<td>1.033</td>
<td>1.16</td>
<td>0</td>
<td>5.91</td>
</tr>
<tr>
<td></td>
<td>Motor-Vehicle</td>
<td>0.085</td>
<td>0.084</td>
<td>0.005</td>
<td>0.552</td>
</tr>
<tr>
<td>Non-Signalized</td>
<td>Cyclist</td>
<td>0.024</td>
<td>0.028</td>
<td>0.001</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td>Pedestrian</td>
<td>0.009</td>
<td>0.010</td>
<td>0.00007</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>Motor-Vehicle</td>
<td>0.029</td>
<td>0.023</td>
<td>0.003</td>
<td>0.176</td>
</tr>
</tbody>
</table>

*Expected number of injuries (obtained from the models) per million cyclists, pedestrians or motor-vehicle occupants per year.*
For all three modes, the risk at non-signalized is much lower than at signalized intersections, as shown in Table 3-3. At signalized intersections, the average risk is greater for cyclists and pedestrians than for motor-vehicle occupants. Figure 3-4 shows that for motor-vehicles, 94% of the intersections sampled have a risk lower than 0.5, whereas for cyclists and pedestrians 89% and 73% of the intersections have a risk greater than 0.5, respectively. At signalized intersections the maximum risk for cyclists and pedestrians are both 11 times greater than for motor-vehicles. At non-signalized intersections the risk for pedestrians is much lower than for cyclists and motor-vehicle occupants.

![Figure 3-4 Distribution of Intersections According to Risk](image)

### 3.6.3. Risk as a Function of Motor-Vehicle Flow

Figure 3-5 shows box plots of cyclist, pedestrian and motor-vehicle risk at signalized (left) and non-signalized (right) intersections according to motor-vehicle volumes. At signalized intersections (Figure 3-5 left) we see that there is no apparent variation in motor-vehicle risk, it remains relatively constant (and low) with respect to all levels of motor-vehicle flow. Cyclist and pedestrian risk on the other hand increase with increasing motor-vehicle flow. This emphasizes once again the danger that motor-vehicles bring to non-motorized modes of transportation. At non-signalized intersections (Figure 3-5 right), this association between motor-vehicle flow and cyclist and pedestrian risk is not as noticeable. This may be because for all modes, the flows and injuries are much smaller and there are many zeros for both at non-signalized intersections.
3.7. CONCLUSION AND FUTURE WORK

This study presents a multimodal safety analysis at signalized and non-signalized intersections. The crash contributing factors are identified for three road users: cyclists, pedestrians and motor-vehicles.

This study considers a sample of 647 signalized and 435 non-signalized intersections on the island of Montreal. The proposed multimodal approach for traffic safety simultaneously analyzes the safety and traffic flow outcomes for motorized and non-motorized traffic. Safety performance functions for the different road users, cyclists, pedestrians and motor-vehicles were developed for both signalized and non-signalized intersections. At signalized intersections, in addition to traffic flows, other important factors associated with injury occurrence were identified: for cyclists, total crosswalk length, presence of a raised median and bus stops; for pedestrians, the number of lanes, presence of pedestrian exclusive signal phasing (all-red or half-red) and commercial entrances and exits; for motor-vehicle occupants, the presence of bus stops and three approaches instead of four. At non-signalized intersections, in addition to flows, the total number of lanes increases cyclist injury occurrence. The objectives were met through the use of an extensive, rich and unique dataset of a large sample of intersections. The applicability of this framework for comparing injuries and risk was also demonstrated.

Among the main findings, we can mention the importance of motor-vehicle traffic, both in total numbers and specific movements (left turn, right turn and through movements) on the injury occurrence for all users and at both facility types. The risk that vehicular traffic imposes on cyclist
and pedestrian injury occurrence, was further emphasized. This highlights the need for safety improvements for cyclists and pedestrians who are, on average, at 14 and 12 times greater risk than motorists, respectively, at signalized intersections. We also confirm that the number of injuries and risk for all modes are greater at signalized than at non-signalized intersections. This work also quantified the effect of crosswalk length and number of lanes on cyclist and pedestrian injury occurrence.

As part of future work, the multimodal analysis approach will be improved by simultaneously modeling cyclist, pedestrian and motor-vehicle injury occurrence and flows in a single modeling framework. Also, this work will be extended to further investigate the effect of correlation among injury outcomes. Another important issue to address is whether or not intersections that are dangerous for cyclists are also dangerous for other modes? If so, this can ease the task of selecting sites for interventions. The hotspot approach, which is widely used in traffic engineering, only targets sites with the highest risk of crash and injury. For a widespread problem such as pedestrian and cyclist safety, this high-risk approach may not result in significant safety improvements for the whole population (Morency and Cloutier, 2006). Another approach comes from public health and has only recently begun gaining ground in traffic engineering. The population approach, which focuses on the large number of people exposed to an average or small risk, instead of the few people exposed to high risk, may provide greater public health benefits (Rose, 1985). For example, interventions may address exposure to moving vehicles and aim at a global reduction in the amount of kilometers driven in an area (Fuller and Morency, 2013). An extension of this paper will investigate the spatial distribution of multimodal injury risk and discuss the appropriateness of both approaches.

Some of the limitations of this study include, the fact that the sample of signalized intersections was not selected representatively, but rather based on the intersections for which manual counts are available. The current sample of non-signalized intersections is representative, however a larger sample would make it possible to further study the effects of geometric design and built environment characteristics on injury occurrence at this type of facility. Also, the analysis in this study should be replicated using police report data as well as a larger sample of intersections in order to validate the results.
3.8. ADDENDUM

The versions of this paper presented at the 24th Canadian Multidisciplinary Road Safety Conference, 2014 and at the Transportation Research Board 93rd Annual Meeting, are different from the paper accepted for publication in Accident Analysis and Prevention. Some of the results that were omitted from the published version of the paper are shown in Appendix B - Additional Multimodal Results.

3.9. REFERENCES


Link between Chapter 3 and Chapter 4

The first two chapters of this thesis focused on modeling injuries and risk simultaneous with bicycle activity and shifted from considering one mode, cyclists (Chapter 2) to considering cyclists, pedestrians and motor-vehicle occupants, applying a multimodal approach (Chapter 3). One limitation of the previous chapters is that only a sample of intersections, for which count and geometric design and built environment data is available or has been collected, were included in the safety and activity analyses (for network screening or for identifying injury factors). To overcome this limitation, the next chapter expands on the previous safety work by developing and proposing a new methodology to map bicycle volumes (activity), injuries and risk for the entire population of sites (intersections and road segments) in a given network. This methodology integrates and combines all the available short- and long-term count data (the data used in the previous chapters) with a new source of bicycle count data - GPS data from a Smartphone application. With volume, injury and risk indicators for the entire population of intersections and road segments in the network, one is able to generate a complete safety portrait of the city for network screening and for analyzing and identifying the contributing factors. From this chapter onward, the cyclist GPS data from a Smartphone application is integrated into the definition of bicycle activity, is used to develop surrogate safety indicators and to compute cyclist speeds and delays and all can be mapped throughout the network. This work is expected to add value to new data collection methods based on Smartphone applications and emerging technologies. Also, this work is the first to carry out in-depth analysis using the Smartphone GPS trip data for cyclists for safety and mobility applications at the network level.

Since the cyclist, pedestrian and vehicle counts, used in the previous chapters, were carried out at intersections, the previous chapters only modeled injury occurrence at a sample of intersections, both signalized and non-signalized. Now, with a data fusion approach, the proposed methodology in the next chapters allows us to go from site-level to network-level safety and mobility analyses.
Chapter 4

Mapping Cyclist Activity and Injury Risk in a Network Combining Smartphone GPS Data and Bicycle Counts
Chapter 4: Mapping Cyclist Activity and Injury Risk in a Network Combining Smartphone GPS Data and Bicycle Counts

Jillian Strauss\textsuperscript{a}, Luis Miranda-Moreno\textsuperscript{a} and Patrick Morency\textsuperscript{b}

\textsuperscript{a}Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, Canada
\textsuperscript{b}Montreal Department of Public Health, Montreal Health and Social Service, Montreal, Canada

This work was presented at the Transportation Research Board 94\textsuperscript{th} Annual Meeting in Washington, D.C., 2015 and at the Canadian Association of Road Safety Professionals (CARSP) conference in Ottawa, Ontario, 2015 where it was awarded the second place paper by the Insurance Bureau of Canada (IBC). This paper has also been published in Accident Analysis and Prevention.

4.1. ABSTRACT

In recent years, the modal share of cycling has been growing in North American cities. With the increase in cycling, the need of bicycle infrastructure and road safety concerns have also been raised. Bicycle flows are an essential component in safety analysis. The main objective of this work is to propose a methodology to estimate and map bicycle volumes and cyclist injury risk throughout the entire network of road segments and intersections on the island of Montreal, achieved by combining Smartphone GPS traces and count data. In recent years, methods have been proposed to estimate average annual daily bicycle (AADB) volumes and injury risk estimates at both the intersection and segment levels using bicycle counts. However, these works have been limited to small samples of locations for which count data is available. In this work, a methodology is proposed to combine short- and long-term bicycle counts with GPS data to estimate AADB volumes along segments and intersections in the entire network. As part of the validation process, correlation is observed between AADB values obtained from GPS data and AADB values from count data, with R-squared values of 0.7 for signalized intersections, 0.58 for non-signalized intersections and between 0.48 and 0.76 for segments with and without bicycle infrastructure. The methodology is also validated through the calibration of safety performance functions using both
sources of AADB estimates, from counts and from GPS data. Using the validated AADB estimates, the factors associated with injury risk were identified using data from the entire population of intersections and segments throughout Montreal. Bayesian injury risk maps are then generated and the concentrations of expected injuries and risk at signalized intersections are identified. Signalized intersections, which are often located at the intersection of major arterials, witness 4 times more injuries and 2.5 times greater risk than non-signalized intersections. A similar observation can be made for arterials which not only have a higher concentration of injuries but also injury rates (risk). On average, streets with cycle tracks have a greater concentration of injuries due to greater bicycle volumes, however, and in accordance with recent works, the individual risk per cyclist is lower, justifying the benefits of cycle tracks.

4.2. INTRODUCTION

Growing climate change as well as energy and health concerns are causing a shift away from motor-vehicle dominance towards “greener” and healthier forms of transportation. Accordingly, a higher modal share of cycling has been observed in recent years in North American cities (Pucher et al., 2011a). With the increase in cycling, the need for bicycle infrastructure and road safety concerns have also been raised in urban areas.

In this context, local governments are seeking to invest more in bicycle infrastructure as well as countermeasures to make roads safer for cyclists, such as the installation of bicycle facilities along busy roads (bicycle paths or cycle tracks), traffic calming measures (bicycle boulevards), bicycle traffic signals, pavement markings (bicycle boxes) as well as others. In the planning of bicycle infrastructure investments and safety countermeasures, new data needs have emerged, such as the accurate estimation of average annual daily bicycle volumes (AADB) at each location of interest or in the entire road network. ADB is essential in many tasks, in particular in the estimation of bicycle injury risk (Strauss et al., 2014, 2013b). Given the importance of bicycle volumes (also referred to as bicycle activity and volume), in recent years, research has been carried out to map bicycle activity to identify injury risk factors and map risk in the network, also referred to as road network screening (Strauss and Miranda-Moreno, 2013). Bicycle flow on each facility and network element is therefore an essential component in many planning tasks such as the identification of corridors with high bicycle activity where infrastructure can be justified, the identification of dangerous routes or intersections where safety countermeasures are required, the
identification of risk contributing factors, air quality studies and many more. Knowing where cyclists are riding and in what numbers serves as a planning tool for cities. Especially, knowing the routes taken by cyclists can reveal whether or not they are following the shortest path or veering off and preferring a slightly longer trip along bicycle infrastructure or on low-traffic streets. This can tell the cities whether or not their bicycle infrastructure is being used and to what extent. In the literature, recent works have proposed methods to estimate AADB by combining short-term (manual) counts in a relatively large sample of sites with long-term (automatic) counts in a small sample of sites (Esawey et al., 2013; Nordback et al., 2013a; Nosal, 2014; Roll, 2013). These methods derive extrapolation factors from the long-term counts and then apply them to the short-term counts. In this approach, AADB is estimated in the large sample of sites with manual bicycle count data. This approach, while useful, only provides a partial and limited picture of the bicycle activity in a network. Manual counts are typically available at a few hundred points in a network. It is practically impossible to have manual or automatic counts at each intersection and road segment in the network so AADB estimates are limited to sites with count data. To our knowledge, no studies have proposed methods to estimate AADB over an entire city's road network including all its signalized and non-signalized intersections as well as its road segments. In an attempt to predict bicycle flows at the entire network level from point locations with count data, some studies have generated land-use regression models (Griswold et al., 2011; Jones et al., 2010; Strauss and Miranda-Moreno, 2013).

The objectives of this work are, to: i) estimate bicycle volumes (AADB) combining Smartphone GPS traces with short-term and long-term counts and ii) map AADB and injury risk throughout the entire Montreal network of segments and intersections. For this purpose, a large sample of bicycle GPS trip data is used. GPS trip volumes at intersections and segments are correlated with average annual daily bicycle volume (AADB) obtained through the combination of manual short-term and automatic long-term counts. Bayesian methods are then applied to obtain injury risk and these results can be mapped for all elements of the road and intersection network for the entire island of Montreal.
4.3. LITERATURE REVIEW

Given the increasing popularity of the bicycle as a mode of transportation and the safety concerns that come along with it, knowledge of where cyclists are riding and in what numbers is vital information for current and future infrastructure planning. As a result, the literature on data collection methods and methodologies to estimate AADB have begun to emerge. In the recent literature, studies have proposed methods to estimate AADB through the combination of short-term (manual) and long-term (automatic) counts (Esawey et al., 2013; Nordback et al., 2013a; Nosal, 2014; Roll, 2013). In a recent report, Nordback et al. (2013a) used this method to pursue the development of models to estimate bicycle volumes based on short-term count data in Colorado. As emphasized by Nordback et al. (2013a), ideally all sites should obtain continuous counts for long periods of time but this is very expensive and not practical, hence the approach that is widely used based on combining available count data. In the city of Davis, the importance of automatic long-term counts has been demonstrated when combined with short-term counts (typically collected using a manual procedure), in order to estimate AADB (Proulx, 2013).

In recent years, technologies to obtain automatic counts have emerged. Some recent reports have evaluated the performance of various technologies. A summary of the available technologies for both short- and long-term counts and how they work is provided by Nordback et al. (2013a) and by Proulx (2013). Nordback et al. (2013b) and a FWHA report (AMEC E&I and Sprinkle Consulting, 2011) also evaluate and discuss the performance and sources of error of these technologies and the quality of the data provided. These technologies include video counts, active infrared, passive infrared, inductive loops and pneumatic tubes.

Another important piece of information for bicycle planning is cyclist route data. This data can be collected by equipping a sample of cyclists with GPS devices. This method usually involves cyclists who are told to ride along specific routes for the purpose of the study. Smartphone applications on the other hand can be used by any and all cyclists possessing a Smartphone, providing the potential for a much larger sample of cyclists and a much wider variety of routes to study. The use of Smartphone applications to collect trip (route) data has gained momentum since the development and implementation of CycleTracks in 2010 by the San Francisco Municipal Transportation Agency (SFMTA). This Smartphone application enables the collection of cyclist trip data to gauge the current bicycle infrastructure and help guide future planning. The SFMTA used this application to develop travel demand models to study where cyclists are riding and how
bicycle infrastructure affects their route choices. Some results have been published and identify that cyclists prefer bicycle lanes over other facility types and are discouraged by steep slopes, turning and long distances (Hood et al., 2011). Following San Francisco’s success, other applications have been implemented in several cities in the United States. For instance, Atlanta launched *Cycle Atlanta*, developed at Georgia Tech, in hopes of collecting data to aid in improving cycling conditions and developing a connected network of bicycle facilities. This application even enables cyclists to comment on trouble areas along their trip. Also, researchers tested San Francisco’s *CycleTracks* application in Austin as a case study (Hudson et al., 2012). During a six-month period between May and October, 2011, over 3,600 routes were recorded and about 300 cyclists provided some information about themselves and their trips such as trip purpose, age, gender and biking frequency. This report presents a methodology to process and analyze cyclist GPS trip data.

The importance and use of AADB has been widely documented in the cyclist safety literature. For instance, we can refer to Brüde and Larsson (1993), Elvik (2009), Lusk et al., (2013), Strauss et al. (2013a) and Strauss et al. (2013c). These studies have identified and confirmed that more cyclists at an intersection translates into more cyclist injuries but lower injury rates due to the non-linear association between bicycle volume and injury occurrence. Bicycle flows serve as an input in safety performance functions and for computing cyclist risk of injury along the facilities of interest. Overall these safety works emphasize the importance of reliable count data for a variety of safety applications.

Overall, previous studies have obtained and applied counts to estimate AADB in a sample of locations. However, there is a lack of methods for estimating AADB and mapping injury risk in an entire network. While GPS data has been used to develop bicycle demand models and identify preferences, it has not been used to achieve this specific and important goal.

There is no doubt that bicycle flow data is a vital component in expanding and planning current and future infrastructure projects as well as evaluating the safety of these projects after implementation. However, few studies, if any, have benefited from this new emerging Smartphone application data collection method to evaluate its value in estimating bicycle exposure measures capable of modeling flows through intersections and along segments providing data for the entire network. Traditionally, studies have relied on short-term counts from a small sample of sites and some have expanded these using long-term counts. Few studies have pursued the estimation of
bicycle flows for an entire network using land-use based regression approaches (Griswold et al., 2011; Strauss and Miranda-Moreno, 2013). Despite these efforts, the previously proposed methods to obtain AADB still require counts from a large number of sites.

4.4. METHODOLOGY

The methodology has several steps:

1. Pre-definition of the network nodes and links.
2. Assign the GPS traces to the network elements (segments and intersections).
3. Obtain AADB volumes from manual short-term and automatic long-term counts and develop an extrapolation function for the GPS data.
4. Validate the predicted AADB from GPS data through the development of Safety Performance Functions (SPF).
5. Apply the predicted AADB for segments and intersections for safety applications.

The above steps are shown in Figure 4-1 and are explained in detail as follows:

![Figure 4-1 Flowchart of Methodology](image)

Figure 4-1 Flowchart of Methodology
4.4.1. Assign GPS Traces to the Network

From the Smartphone GPS traces, the first step after mapping all the raw GPS observations (x,y) is to assign each observation to the network elements, to a road segment or to an intersection. For this purpose, a buffer approach is used. To obtain the number of GPS trips traveling along each road segment and passing through each intersection on the island of Montreal, we can reduce the number of steps and processing time by connecting all the points from the same trip to form lines. In other words, the thousands of GPS points for each trip are reduced to one line per trip. Due to the noise of the GPS data as a result of buildings, trees and adverse weather as well as the accuracy of the Smartphone’s GPS, to capture all the trips, 35 metre buffers were created around each road segment and each intersection. A small portion of the study area is shown in Figure 4-2 to show these buffers. Also, note that there is no overlap between the segment and intersection buffers. One starts where the other one ends and vice versa. The GPS trip lines were then intersected with these buffers and the number of lines (representing the trips) were counted along each segment and passing through each intersection.

![Figure 4-2 Segment and Intersection Buffers and GPS Points](image)

4.4.2. AADB Volumes from Counts and Extrapolation Functions

Consider a network for which manual short-term counts are available for a specific number (sample) of locations (intersections and/or segments). Also, assume that there are automatic
counters installed at a few locations from which long-term counts are available. In order to standardize short-term counts and obtain AADB, the following steps are required:

1. Compute hourly, daily and monthly expansion factors from the long-term count locations using one of the factoring methods proposed by Nosal and Miranda-Moreno (2014) and Strauss (2012a).

2. Match a long-term counter to each short-term count location based on several criteria (ridership patterns, proximity, land use, type of facility, etc.).

3. Apply the developed expansion factors in order to obtain the average annual daily bicycle volumes (AADB) at all locations with short-term counts.

4. For each segment and intersection in the network, sum the number of GPS trips associated with each. Then, develop a function that associates the number of GPS trips with the AADB volumes obtained in the previous step. The aim of this extrapolation function is to find the constants (parameters) that convert the GPS trips from the Smartphone application to AADB estimates. This function also identifies factors that are correlated with AADB which helps to correct for the bias in the data. This function is of the form:

   \[ \text{AADB}_{ik} = \beta_k \cdot T_{ik} + \alpha_k \quad \text{with} \quad \alpha_k = f(x_i; \gamma) \]

Where,
- \( i \) stands for network element \( i \) of type \( k \) (signalized intersection, non-signalized intersection or segment).
- \( \text{AADB}_{ik} \) stands for average annual daily bicycle volume at location \( i \) of type \( k \) derived from the combination of short- and long-term counts.
- \( \beta_k \) stands for the parameter weighing the average number of GPS trips per day during the study period denoted as \( T_{ik} \).
- \( \alpha \) stands for the correction factor associated with geometric design and built environment characteristics denoted by \( x \), such as the presence of a cycle track or bicycle path, as well as the distance to downtown.
- \( \gamma \) is the regression parameter associated with each variable.

Linear, non-linear and count data regression models are tested to estimate the regression parameters, \( \beta \) and \( \alpha \). The main criteria for variable and function selection are goodness of fit
measures, which depend on the model formulation. Separate $\beta$ and $\alpha$ regression parameters are also obtained for the different network elements, $k$.

### 4.4.3. Validation of Predicted AADB

There is a need to validate the extrapolation factors developed from the GPS trip data. One way to validate the predicted AADB values is to develop safety performance functions (SPF) based on AADB values obtained from the GPS data and compare the parameter coefficients with the SPFs previously developed based on short- and long-term count data.

### 4.4.4. Safety Applications

After calibrating the AADB function, it can be applied to all intersections and segments in the network and AADB can be predicted for the entire network. Cyclist injuries can then be overlaid and the number of injuries can be combined with AADB to obtain a measure of cyclist risk at each intersection and segment. Raw injury rates (risk) at intersections and along segments can then be mapped throughout the entire network. A better way to compute injury risk is using statistical models and then computing empirical Bayes (EB), this way combining both the observed and predicted injuries to obtain an even better estimate than each separate piece of information. The EB estimate of cyclist risk can be obtained and mapped throughout the entire network. To obtain the EB estimator, a Poisson/Gamma model is assumed, also referred to as Negative Binomial model. The EB estimator is then computed as shown in Equation 1.

$$E[\theta_{i|y}] = (1 - w_i)y_i + w_i\mu_i$$

(1)

Where, $w_i = \frac{\phi}{\mu_i + \phi}$, $\phi$ is fixed in the model as $1/\alpha$ ($\alpha$ is the dispersion parameter from the Negative Binomial SPF model). $y_i$ represents the observed cyclist injuries at intersection $i$ and $\mu_i$ is the expected number of cyclist injuries obtained by applying the model to the data.

We need to generate EB for the entire population of intersections and segments but motor-vehicle flows as well as many geometric design and built environment characteristics have been collected for a sample of sites and are not available for the entire population of intersections and segments. For this reason, new SPF functions, with potentially different variables, are required and generated in this work. With these models, EB for injuries are obtained, and by accounting for the predicted AADB, we can obtain a measure of risk (injury rate) per cyclist throughout the entire network.
network of intersections and segments. For segments, the risk also accounts for the length of the segment so the risk has units of injury rate per cyclist per kilometre. Risk, denoted by $\overline{R}_i$, can be expressed as shown in Equation 2.

$$\overline{R}_i = \overline{\theta}_i \times 10^6 / 365 \times t_i \times Z_i \quad (2)$$

Risk is based on the posterior mean of injury frequency at intersection $i$, $\overline{\theta}_i$, and is expressed as the posterior injury rate per million cyclists, $Z_i$, per unit of time, $t_i$.

Using the results for flows, injuries and risk at intersections, we are able to identify intersections or areas where for example, flows are high but there are no bicycle facilities nearby. More importantly, we are able to explore where cyclist injury risk is high and there are no bicycle facilities and identify hotspots.

4.5. DATA

The data sources are GPS traces, short- and long-term counts, injury, geometric design and built environment data.

4.5.1. Smartphone GPS Trips and Traces

This work uses GPS cyclist trip data collected from the Mon RésoVélo Smartphone application for both android and iOS Smartphones (Jackson et al., 2014). When cyclists begin their trip, they start this application which records their trips and provides second-by-second latitudes, longitudes and timestamps (depending on the phone and the quality of the GPS signal, these may be recorded less often). This work uses data collected from July 2nd, 2013, which is the day the application was launched, until November 15th, 2013, which represents the day when most of the bicycle facilities in Montreal close for the winter. The study period is therefore 137 days long and in this short time, over 10,000 trips were recorded by almost 1,000 cyclists, representing over 16 million GPS data points needing to be processed to meet our objectives.

4.5.2. Short-Term and Long-Term Counts

In 2008 and in 2009, the city of Montreal carried out 8-hour manual cyclist counts at over 600 signalized intersections on the island of Montreal. In the summer and fall seasons of 2012, we set
out to collect count data for 435 randomly selected non-signalized intersections throughout the island. These samples of intersections are shown in Figure 4-3.

Also, in Montreal, inductive loop and pneumatic tube counters have been installed in 30 different locations. Some of these counters have been installed since 2008 while others have been installed in more recent years. Counters on Rachel, de Brebeuf, Cote Sainte Catherine, Saint Urbain, Berri and two separate ones on de Maisonneuve were installed and collected data for at least one complete year covering all months of the year and not just the cycling season. The remaining counters were installed anywhere from one to eight months. These long-term counts have been used to develop hourly, daily and monthly expansion factors to adjust the manual short-term counts to AADB values based on the hour, day and month that each manual count was taken. It is important to mention that the short-term counts and trips are not taken from the same year. This can affect the accuracy of the estimates.

Figure 4-3 Intersections with Manual Counts and Automatic Count Locations
4.5.3. Cyclist Injuries

This study focuses on cyclist injuries which occurred at intersections and along road segments over a six-year study period from 2003 to 2008 obtained from ambulance services. Figure 4-4 shows the distribution of raw cyclist injury data. Accidents are considered as having occurred at an intersection if they are within 15 metres of the centre point of the intersection and are otherwise assigned to road segments. During this time period, over 5,000 cyclists were injured at intersections and over 3,500 were injured along segments. Although ambulance data may be biased towards more severe injuries, in Montreal, this source of data identified more cyclist injuries than police reports. A crash involving two cyclists would be considered as two separate injuries, however this situation arises in a very small percentage of cases in this dataset.

4.6. RESULTS

After defining the intersections and segments in the network and assigning the GPS traces to these network elements, the extrapolation function to obtain AADB from GPS traces is developed, leading to the estimation of AADB, injuries and risk throughout these network elements for the entire island of Montreal.
4.6.1. AADB Estimation

The linear models for predicting AADB from GPS data provide a better fit than the Poisson models and therefore linear model formulation was selected. Models were generated separately for signalized intersections, non-signalized intersections and segments. Models were selected based on the highest R-squared value. Also, only variables significant to the 95% level and with a correlation value of less than 0.4 between each variable, were selected in the models.

In order to capture the effects of bicycle facilities on bicycle flow at intersections, variables were generated combining the flow and facility type: GPS count-No facilities, GPS count-Bicycle path and GPS count-Cycle track.

The results for the model developed for signalized intersections are shown in Table 4-1a. This model has an R-squared value of 0.70. According to this model, the resulting predicted AADB flows are always greater than zero, even when the GPS count is zero. This is achieved by forcing the “distance to downtown” variable to be a value of 15 kilometres for all distances both equal to and greater than 15 kilometres. The results identify that the presence of bicycle facilities have an increasing effect on cyclist numbers reflected by the parameters of “GPS count-Bicycle path” and “GPS count-Cycle track”. More specifically, the results show that the presence of a cycle track has the greatest effect on increasing cyclist numbers, more than both no facilities and bicycle paths. Also, as expected, these results show that the farther away you are from the central neighbourhoods of the city, bicycle flows decrease. AADB values at signalized intersections are then predicted using this model.

Similarly, a model was generated for non-signalized intersections as shown in Table 4-1b. In general, bicycle flows are much lower at non-signalized than at signalized intersections as is reflected in the lower parameter estimates. These results show the same effect of bicycle facilities on flows and again show that the presence of a cycle track has the greatest effect on increasing bicycle flows and bicycle flows decrease as you get farther from the central neighbourhoods. The R-squared value of this model is 0.58.

For segments, a different approach was used to best capture the effects of the presence of bicycle facilities on bicycle flow. Three models were developed, for segments with: 1) cycle tracks, 2) bicycle paths and 3) no bicycle infrastructure. The results are shown in Table 4-1c. It is worth mentioning that in the “no facility” model, the constant has been suppressed. This is to reflect the...
unlikelihood of there being any bicycle flow along segments without any facilities that did not witness any GPS traces during the study period.

After applying these models and obtaining the AADB for all intersections and segments, AADB can be mapped for the entire network. The predicted AADB values are shown in Figure 4-5.

Table 4-1 AADB Model Results

a) Signalized Intersections

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS count - No facilities</td>
<td>1575.5</td>
<td>83.6</td>
<td>0.000</td>
</tr>
<tr>
<td>GPS count - Bicycle path</td>
<td>919.3</td>
<td>109.6</td>
<td>0.000</td>
</tr>
<tr>
<td>GPS count - Cycle track</td>
<td>2387.9</td>
<td>79.5</td>
<td>0.000</td>
</tr>
<tr>
<td>Distance to downtown*</td>
<td>-15.34</td>
<td>3.9</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>238.4</td>
<td>35.2</td>
<td>0.000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>638</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*All observations with distance greater than 15km, forced to be 15km

b) Non-Signalized Intersections

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS count - No facilities</td>
<td>175.4</td>
<td>12.3</td>
<td>0.000</td>
</tr>
<tr>
<td>GPS count - Bicycle path</td>
<td>157.6</td>
<td>15.1</td>
<td>0.000</td>
</tr>
<tr>
<td>GPS count - Cycle track</td>
<td>567.2</td>
<td>56.2</td>
<td>0.000</td>
</tr>
<tr>
<td>Distance to downtown*</td>
<td>-24.1</td>
<td>2.66</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>378.4</td>
<td>29.8</td>
<td>0.000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>438</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*All observations with distance greater than 15km, forced to be 15km

c) Segments

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cycle Track</th>
<th></th>
<th>Bicycle Lane</th>
<th></th>
<th>No Facility</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS flow</td>
<td>2753.7</td>
<td>316.5</td>
<td>0.000</td>
<td>1287.8</td>
<td>200.0</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>1557.1</td>
<td>164.7</td>
<td>0.000</td>
<td>1387.1</td>
<td>121.0</td>
<td>0.000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.52</td>
<td></td>
<td></td>
<td>0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>70</td>
<td>14</td>
<td></td>
<td>36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- In the no facility model the constant was suppressed
For validation purposes, intersections with both sources of data, manual short-term counts and GPS traces, were extracted and injury occurrence models were developed with AADB from counts and AADB from GPS traces. This was done for signalized intersections and still remains to be validated for non-signalized intersections and segments. The results are shown in Table 4-2. These results show that the regression parameters estimated from manual counts are very similar to those estimated with GPS trips. This further validates the GPS trip data as a reliable source of bicycle flow data to be used in a variety of safety analyses.

**Table 4-2 Injury Model Results to Compare AADB from Manual Counts to AADB from GPS Trips for Validation Purposes**

<table>
<thead>
<tr>
<th>Variable</th>
<th>AADB from manual counts</th>
<th>AADB from GPS trips</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
</tr>
<tr>
<td>Ln* bicycle flow (AADB)</td>
<td>0.510</td>
<td>0.054</td>
</tr>
<tr>
<td>Ln* right turn motor-vehicle flow</td>
<td>0.174</td>
<td>0.066</td>
</tr>
<tr>
<td>Ln* left turn motor-vehicle flow</td>
<td>0.138</td>
<td>0.055</td>
</tr>
<tr>
<td>Crosswalk width</td>
<td>0.010</td>
<td>0.003</td>
</tr>
<tr>
<td>Bus stop</td>
<td>0.468</td>
<td>0.154</td>
</tr>
<tr>
<td>Raised median</td>
<td>-0.478</td>
<td>0.154</td>
</tr>
<tr>
<td>Constant</td>
<td>-6.53</td>
<td>0.627</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-621.1</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>1258.1</td>
<td></td>
</tr>
<tr>
<td>Dispersion parameter</td>
<td>0.553</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>635</td>
<td></td>
</tr>
</tbody>
</table>

*Ln = natural logarithm*
4.6.2. Risk Analysis

Injury risk is estimated at intersections and along segments using the empirical Bayes (EB) estimator, which is a popular technique in road safety analysis. This technique combines both the observed and predicted injuries to obtain a better risk estimate than observed injuries or raw rates. The injury model results shown in Table 4-2 were computed based on the sample of intersections with inventory data (motor-vehicle flows and site-specific geometric design and built environment characteristics). For safety applications, flows, injuries and risk are vital information required for all (entire population) intersections and segments. Since we generated a model to predict AADB from GPS data, we have AADB for all intersections and segments but we only have motor-vehicle flows for the sample of sites. Proxy variables such as road type can be used to account for motor-vehicle flow. Also other variables for which we have data for all sites can be tested in the models, such as the presence and number of bus stops, number of approaches and the neighbourhood or distance to downtown. In order to obtain EB for all intersections and all segments, new safety performance functions were developed using the variables available for all intersections and segments in the network. Three models were developed, for: 1) signalized intersections, 2) non-signalized intersections and 3) segments. The results are shown in Table 4-3.

The predicted AADB values for signalized intersections, non-signalized intersections and segments are shown in Figure 4-6. Using the predicted injuries obtained from the models, EB for injuries can be obtained. By combining these EB injuries with AADB flows, we are able to compute EB of injury rates (risk) per cyclist along road segments and at intersections throughout the entire network. EB risk at all facility types is also shown in Figure 4-6. This figure identifies that for both signalized and non-signalized intersections, AADB is greater in the central neighborhoods of the island. Risk follows the opposite pattern of injuries and flow, and cyclist risk is greater outside the central neighbourhoods. In other words, more cyclists implies more cyclist injuries however the association between flows and risk is negative and non-linear. The maps for segments do not seem to follow the same trend as for intersections. This is related to the fact that overall there are so few cyclist injuries along segments but also because all injuries along segments occurred at only 4% of all segments in the network. This means that 96% of segments (over 42,000) did not witness any cyclist injuries during the study period.

Further, combining risk with bicycle infrastructure (cycle tracks and bicycle paths), we can identify areas where cyclist risk of injury is high and which do not have any bicycle facilities, as
well as identify areas with high flows and still no facilities. This can aid the city in showing where risk through intersections and segments is high. These results can serve as a planning tool for the city of Montreal for the development and expansion of bicycle infrastructure throughout the city.

Based on injury risk analysis results, we are able to compare risk at signalized and non-signalized intersections as well as between intersections and segments with and without arterials and cycle tracks. Figure 4-7 provides box plots showing the distributions of EB injuries and EB risk for intersections and compared for different geometric designs. Figure 4-7a shows that on average, signalized intersections are 4 times more dangerous in terms of injuries (mean value of 0.78) than non-signalized intersections (mean value of 0.18). In terms of risk, signalized intersections are, on average, 2.5 times more dangerous. Signalized and non-signalized intersections can then be split further into those with and without arterials. 53% of signalized intersections have at least one arterial whereas only 22% of non-signalized intersections have an arterial. At signalized intersections, the mean number of injuries as well as the mean risk of injury have the same value for intersections both with and without arterials (Figure 4-7b). On average, at non-signalized intersections, both injuries and risk are almost 2 times greater at intersections with arterials compared to those without (Figure 4-7c). This highlights the importance of cyclist safety on arterials.

The intersections can also be split into those with and without cycle tracks. 11% of signalized and 7% of non-signalized intersections have cycle tracks. For signalized intersections, as expected, since flows are in general greater along cycle tracks compared to locations without cycle tracks, injury occurrence is over 2 times greater at intersections with cycle tracks (Figure 4-7d). However, when flows are accounted for in the risk measure, we see that risk is about 2 times lower at intersections with cycle tracks. At non-signalized intersections, injuries are greater with cycle tracks but the risk is very similar (Figure 4-7e).
## Table 4-3 Injury Model Results for Signalized Intersections, Non-Signalized Intersections and Segments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Signalized Intersections</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln* bicycle flow (AADB)</td>
<td>0.330</td>
<td>0.017</td>
<td>0.000</td>
</tr>
<tr>
<td>Bus stop</td>
<td>0.413</td>
<td>0.081</td>
<td>0.000</td>
</tr>
<tr>
<td>Three approaches</td>
<td>-0.685</td>
<td>0.114</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.57</td>
<td>0.139</td>
<td>0.000</td>
</tr>
<tr>
<td>Alpha</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations (entire population)</td>
<td>2288</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-Signalized Intersections</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln* bicycle flow (AADB)</td>
<td>0.385</td>
<td>0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>Arterial or collector</td>
<td>1.048</td>
<td>0.047</td>
<td>0.000</td>
</tr>
<tr>
<td>Three approaches</td>
<td>-0.913</td>
<td>0.041</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.94</td>
<td>0.070</td>
<td>0.000</td>
</tr>
<tr>
<td>Alpha</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations (entire population)</td>
<td>23819</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Segments</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln* bicycle flow (AADB)</td>
<td>0.336</td>
<td>0.020</td>
<td>0.000</td>
</tr>
<tr>
<td>Arterial or collector</td>
<td>0.684</td>
<td>0.052</td>
<td>0.000</td>
</tr>
<tr>
<td>Downtown boroughs</td>
<td>0.495</td>
<td>0.071</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.99</td>
<td>0.091</td>
<td>0.000</td>
</tr>
<tr>
<td>Alpha</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations (entire population)</td>
<td>44314</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Ln = natural logarithm*
Figure 4-6 AADB and EB Risk at Signalized Intersections (top: a and b), Non-Signalized Intersections (middle: c and d) and Segments (bottom: e and f)
4.7. CONCLUSION AND FUTURE WORK

One of the main limitations in the current literature is the lack of bicycle activity data available at the entire network level for cities. In this work we tackled this issue by exploring the use of
Smartphone GPS cyclist trip data combined with short- and long-term counts for estimating bicycle flows for the entire network. With this data, average annual daily bicycle flows (AADB) were predicted for all intersections and road segments. The main benefit of using GPS data in combination with counts, is that GPS data provides a sample of the volumes over the entire road network (large spatial coverage) and the available counts help to extrapolate these to AADB values. This method then enables us to map bicycle volumes in the network which is vital information in planning processes and safety studies, among other applications.

Applying the proposed methodology, models to predict AADB from GPS data at signalized and non-signalized intersections, as well as at segments were developed. A validation process was implemented for the models used to predict flows at signalized and non-signalized intersections as well as segments. Applying these models made it possible to map bicycle flows throughout the entire road and intersection networks on the island of Montreal. Modeling injuries with these flows as well as with geometric design and built environment attributes, we can obtain a Bayesian prediction of injuries and map these throughout the entire network. Further, the prediction of injuries can be combined with the AADB flows to obtain a measure of cyclist injury risk which can then also be mapped throughout the entire city network. Based on this measure of risk, we are able to identify hotspots and justify segments and intersections in need of treatments.

Among other results, it was found that cyclist injury risk is greater outside the central neighbourhoods of the island where bicycle infrastructure is lacking. Injuries and AADB on the other hand, are highest in these central neighbourhoods. Also, on average, signalized intersections are 4 times more dangerous in terms of injuries than non-signalized intersections. In terms of risk, signalized intersections are, on average, 2.5 times more dangerous. Splitting the intersections into those with at least one street classified as an arterial, reveals that both injuries and risk are greater compared to those without any arterials. Signalized intersections, in particular those with arterials, have both high expected injury occurrence and high injury rates.

Splitting intersections into those with and without cycle tracks, reveals that on average, injury occurrence is greater at both signalized and non-signalized intersections with cycle tracks because cyclist activity is higher at these intersections. In terms of risk, at signalized intersections with cycle tracks, the risk is lower than those without cycle tracks.

For segments, both injuries and risk were found to be highest in the central neighbourhoods but overall, risk is much lower at segments than at intersections. On average, both injuries and risk
are greater along segments classified as arterials. Segments with cycle tracks also have more injuries than those without but the differences in terms of risk are minor.

This work benefited from the richness of GPS data to be able to map flows, injuries and risk for cyclists, throughout the entire network of intersections and road segments. GPS data serves as a vital input in many infrastructure planning endeavors since it provides a large spatial portrait of bicycle routes throughout an entire city and can identify, for example, where flows are high but there is no bicycle infrastructure or where risk is high and still there is no infrastructure. Knowledge of where risk is high and where bicycle flows are also high is vital information for cities and provides insight into where to best build facilities for cyclists and where other infrastructure and designs are needed to supply safe routes for cyclists throughout cities. In this work we were able to map the distribution of injuries and risk at the entire network level for the city of Montreal. This means that the city of Montreal can easily apply any approach to select hotspots, either based on injuries or risk, and to select sites for treatments. Most studies, relying on other bicycle flow data collection methods, are not able to predict the individual risk per cyclist on all facilities for an entire city.

As part of future work, we will develop and apply a more elaborate map-matching technique to assign GPS points and traces to the road network elements. A larger GPS trip dataset, from cycling season 2014, will be added to the current dataset of 2013 and the accuracy of the current models will be evaluated. The lower correlation between AADB measures along segments might be associated with the fact that few point locations with counts are currently available along segments. The models for segments should and will be validated with more data. The same proposed methodology should be implemented and validated using other cities with similar data. Self-reported incidents and accidents could also be collected through the Smartphone application and injury maps compared with those obtained using ambulance or police report data.

4.8. ACKNOWLEDGMENTS

We acknowledge the financial support provided by the FQRNT - *Programme de recherche en sécurité routière FQRNT-MTQ-FRSQ*. We would like to thank the Montreal Department of Public Health and Urgences-santé for collecting and validating the injury data as well as the city of Montreal and Yannick Roy for providing motor-vehicle and bicycle flow data and Charles Chung
from Brisk Synergies as well as the city of Montreal for the GPS data. All remaining errors and the views expressed in this research are, however, solely ours.

4.9. ADDENDUM

After publishing this chapter, the authors implemented a map-matching algorithm, \textit{TrackMatching} (Marchal, 2015), to snap the raw GPS data onto the road network using \textit{OpenStreetMap}.

Also, the version of this paper presented at the Transportation Research Board 93rd Annual Meeting, is different from the paper published in Accident Analysis and Prevention.

Appendix C - Map-Matching, Additional Safety Application and GPS Data, provides the correlation results of counts using the map-matching technique and using the simple buffer approach as well as injury maps and the identification of hotspots.

Also, due to space restrictions of the journals and conferences, the GPS data was not thoroughly explained. A more detailed description of the GPS data is provided in Appendix C.

4.10. REFERENCES


Link between Chapter 4 and Chapter 5

The previous chapter developed a methodology to obtain average annual daily bicycle flows throughout intersections and segments combining short- and long-term counts with GPS data. The results made it possible to map bicycle activity in the entire network. Once these were obtained, we overlaid injuries to compute the individual risk faced by each cyclist and mapped these throughout the network. Chapter 4 relied on reported cyclist injury data to develop safety performance functions and to compute risk in the network. Relying on reported injury data has several limitations. Injury data is based on injuries which occurred and were reported to police, to the hospital or to insurance companies. Often, accidents involving a cyclist and a motor-vehicle, resulting in minimal or no injury to the cyclist, go unreported. Also, when injuries are reported, in many cases, the reports are incomplete and the location of the accident is not properly recorded, resulting in the potential for the data to be unusable in analysis. In addition, injuries (accidents) often need to be recorded over several years in order to obtain enough data for analysis. Most importantly, to obtain injury data, an accident involving a cyclist had to occur and due to the vulnerability of cyclists, accidents involving cyclists often result in serious injuries or fatalities.

In recent years, studies have begun to try to shift from an injury-based to a surrogate-based safety approach, in other words to shift from a reactive to a proactive approach, by defining observable and measurable events that can be related to accidents but do not actually involve accidents or injuries.

In response to the shift to a proactive approach, the next chapter develops and proposes a methodology to extract dangerous deceleration values derived from GPS data as a surrogate safety measure. Hard braking is an evasive action, which is observable and measurable, which may have occurred if a cyclist was about to hit or be hit by a car, or a stationary object or it may identify locations where there are faults in the geometric design of the roads and intersections. The proposed surrogate safety measure is validated by investigating its correlation with observed injuries.
Chapter 5

Cyclist Deceleration Rate as Surrogate Safety Measure in Montreal using Smartphone GPS Data
Chapter 5: Cyclist Deceleration Rate as Surrogate Safety Measure in Montreal using Smartphone GPS Data

Jillian Strauss\textsuperscript{a}, Sohail Zangenehpour\textsuperscript{a}, Luis Miranda-Moreno\textsuperscript{a} and Nicolas Saunier\textsuperscript{b}

\textsuperscript{a}Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, Canada
\textsuperscript{b}Department of civil, geological and mining engineering, Polytechnique Montréal, Montreal, Canada

Submitted for publication in Transportation Research Part F: Traffic Psychology and Behaviour: Special Issue on Road Safety as Reflected by Empirical Non-crash Data.

5.1. ABSTRACT

Bicycle safety is an important issue in the context of sustainable mobility. Many injuries happen every year in urban areas in North American cities with positive trends in bicycle usage. To study safety, and in particular to map injury risk in a network, the traditional safety approach which has been used is based on historical accident data which contains several known limitations. This paper proposes a complementary approach based on a simple surrogate safety measure that occurs more frequently, is observable and can be related to cyclist accidents and safety. The surrogate safety measure is defined based on deceleration rates in order to identify hard breaking situations. This work uses data collected from a GPS Smartphone application for cyclists to develop a methodology to obtain deceleration rate at intersections and segments, to explore the relationship between the number of observed injuries and deceleration rate (DR) and to validate DR as a surrogate safety measure. Using the Spearman rank correlation coefficient, we compared the ranking of sites based on the expected number of injuries using the Empirical Bayes estimator (EB) and based on deceleration rates. Based on Spearman’s coefficient, the ranks of EB injuries and dangerous decelerations have a correlation of 0.55 at signalized intersections, 0.37 at non-signalized intersections and 0.48 at segments. Isolating intersections and segments in the central business district of Montreal, the coefficient stays the same for signalized intersections and increases to 0.55 for non-signalized intersections and to 0.58 for segments. These results show that deceleration rate,
obtained from GPS data, can be used as a surrogate safety measure for cyclists, to identify sites that are dangerous prior to witnessing a high number of injuries at these sites.

5.2. INTRODUCTION

Bicycle usage has reported some positive and increasing trends in many urban areas in North America. While cycling as a mode of transportation has grown in numbers in both Canada and the United States, the level of biking is greater in Canada and overall cycling levels vary across American states and Canadian provinces and territories (Pucher et al., 2011b). As bicycle flows increase, so do concerns for cyclist safety which has become a critical issue in many cities. In Canada, from 2008 to 2012, cyclist injuries represented 1.8% to 2.9% of all fatalities, representing an average of 54 cyclist fatalities per year (Transport Canada, 2014). While these numbers may not seem too big, the modal share of cycling in Canada ranges from 0.2 to 2.6. Now that cyclist numbers are on the rise, cyclist safety has become a major concern. This is the case in Montreal where the modal share of cycling is on average 2% and reaches as high as 10% in the central boroughs of the island during the summer months (Vélo Québec, 2011b). On the island of Montreal from 2005 to 2009, 3 fatalities, 36 major injuries and 651 minor injuries were reported for cyclists (Vélo Québec, 2011a).

According to our past research, the injury risk for a person traveling through an intersection as a cyclist is 14 times higher than an individual traveling in a car as the driver or passenger (Strauss et al., 2014). The majority of safety studies reported in the current literature are based on the traditional approach using historical accident or injury data documented in police reports or from ambulance intervention services. This research is very important for shedding light on the role of traffic volumes, built environment and geometric design features as well as towards the identification of dangerous locations and corridors. This research is based on three critical pieces of information: traffic counts, geometric design and observed injury data. While useful, this approach has some limitations. Several years of injuries are required for the analysis, given the low mean value for the occurrence of cyclist injuries at the locations of interest. Collecting count data requires extensive time and resources, so in most cases, counts were made only at a small number of sites in past studies. Also, several accidents go underreported, especially those with very minor or no injuries.
New sources of data and analysis approaches have begun to emerge in the literature to overcome some of these limitations. New sources of data, such as GPS data collected from Smartphones have emerged. These new sources offer the opportunity to complement the traditional approach and deal with some of its gaps since now count and trip data can be collected in the entire network, showing all the facilities used by cyclists in the city. Based on cyclist speeds and acceleration profiles, surrogate safety measures can be developed and extracted from the GPS data, such as hard braking. A dangerous threshold can be defined and the number of occurrences can be counted. This enables a more proactive approach to identifying dangerous locations or problems with the current built environment and geometric design as opposed to the traditional approach which is reactive. In other words, by identifying surrogate safety measures, the safety of sites can be determined and potentially treated, without having to wait several years for accidents to occur and for damages and harm on the population. Surrogate safety measures can also be used to evaluate the safety of treatments by comparing conflicts prior and post treatment. Furthermore, during this time, traffic and bicycle flows are likely to change which makes the attribution to causes even more challenging.

Surrogate safety measures include: time-to-collision (TTC), post-encroachment-time (PET), gap time (GT) as well as deceleration rate (DR) (Ismail, 2010). TTC, PET and GT generally require video data from which these measures can be computed manually or automatically. For DR, GPS data can be used. In this work, cyclist GPS data from the Mon RésoVélo Smartphone application is used (Ville de Montréal, n.d.).

This work has two main objectives, to: 1) develop a methodology to obtain a surrogate safety measure based on deceleration rate for cyclists at intersections and segments using Smartphone GPS data and, 2) explore the relationship between observed injuries and deceleration rate (DR) and validate DR as a surrogate safety measure.

5.3. LITERATURE REVIEW

Many surrogate safety measures have been developed and tested. These include time-to-collision (TTC), post-encroachment-time (PET), gap time (GT) and deceleration rate (DR). The first surrogate measures of safety developed in the 1960s and 1970s relied on the observation of traffic conflicts by trained observers resulting in many discussions over the validity and objectivity of the data. This has since evolved and the current research explores mostly surrogate safety measures
using video data. Automated methods for surrogate safety analysis for cyclists have begun to emerge in the literature (Kassim et al., 2014; Sakshaug et al., 2010; Sayed et al., 2013; Zangenehpour et al., 2015b). A very recent study in Montreal used video data to study the safety effectiveness of cycle tracks using an automated method to obtain PET as a surrogate measure of the interactions between cyclists and turning vehicles traveling in the same direction (Zangenehpour et al., 2015b). A study in Vancouver presented the use of an automated method to obtain TTC to identify the severity of cyclist interactions at one busy intersection (Sayed et al., 2013). Another recent study in Ottawa evaluated cyclist-vehicle interactions at signalized intersections based on PET (Kassim et al., 2014). A study in China used video data collected at one signalized intersection to study cyclist behaviour, specifically cyclist crossing speed, gap and lag acceptance and group riding behaviour (Ling and Wu, 2004).

Specific to intersections, additional surrogate measures include delay, travel time, approach speed, percent stops, queue length, red-light violations as well as speed and deceleration distributions. According to Balasha et al. (1980), for vehicles, if a vehicle puts on the brakes, this is a sign of conflict and can be used as a surrogate measure of safety. Hard braking is an evasive action which had it not been, may have led to an accident. In this work we will explore the applicability of this same theory to cyclists. A study in Atlanta, Georgia, showed that the frequency of hard braking is correlated with crash occurrence for vehicles (Jun, 2006).

Many studies focusing on speed and acceleration profiles, as well as delays, have used GPS data since these measures are fairly easy to extract from this source of data. One study using GPS data, looked at the braking behaviour of both vehicles and cyclists travelling along a predefined track (Dang et al., 2014). Another study used GPS data to identify design inconsistencies on rural roads and their effects on vehicle speed (Cafiso and Cerni, 2012). A more recent study in Beijing, China, installed drive recorders in 50 taxi cars to record decelerations and model the relationship between conflicts and accidents (Lu et al., 2011). Lusk et al. (2011) developed a model which shows that the average number of accidents increases as the average number of conflicts increases but this relation is not linear and as the number of conflicts increases the rate of increase reduces. A study in Atlanta, Georgia, collected GPS data from drivers who both did and did not experience accidents to see the effect that being involved in an accident has on driving behaviour (Jun, 2006). Another study in Atlanta, Georgia, used the 85th percentile of speed variance collected from GPS
receivers installed in over 400 cars to explore its relationship with crash frequency (Boonsiripant et al., 2007).

Many studies have focused on finding empirical evidence to prove that there is a relationship between surrogate safety measures and observed accidents and one of the main drawbacks of the traffic conflict technique is its reliability and validity (Hauer and Garder, 1986; Hauer, 1982; Williams, 1981). The conflict technique dates as far back as 1967 and still no consensus has been reached concerning whether or not surrogate safety measures are viable alternatives to accidents. The advantages of surrogate measures over accident data are numerous. Conflict data can be collected in as little as several hours whereas waiting for accident data can take years and the effect of environment, road design, traffic signals and signs as well as treatments cannot be evaluated in the short term.

There exists several gaps in the current literature. Previous studies have obtained surrogate measures for a sample of sites. Previous studies based on video data, require that data be collected for many hours at a large number of sites which is not an easy task. Also, extracting the data is not an easy task and requires validation to check the accuracy of the method. The previous work using GPS data equipped a sample of cars or bicycles with equipment to compute decelerations, surrogate measures for cyclists have not been thoroughly developed using GPS data for a large sample of locations. Also using a Smartphone application, dangerous decelerations can be obtained for the entire road network of a large city: this is demonstrated in this paper for the island of Montreal.

5.4. METHODOLOGY

The methodology has several steps. The data was processed using the geographic information system software ESRI ArcMap version 10.1. The steps are shown in Figure 5-1 and described in detail as follows.

1. Apply data cleaning rules
2. Separate intersections and segments
3. Compute instantaneous speeds
4. Compute instantaneous accelerations and decelerations
5. Explore the relationship between injuries and decelerations
5.4.1. Cleaning Rules

Due to the noise of the GPS data as a result of buildings (especially in downtown Montreal where there are many high-rise buildings), trees and adverse weather conditions, as well as the accuracy of the Smartphone GPS and if it malfunctions, the data needs to be checked and cleaned. After thorough inspection of the data, the following cleaning rules have been set up:

**High speed**: Using the instantaneous point-to-point speeds, the average speed for each entire trip can be computed. If the average trip speed is greater than 30 km/h, it is very unlikely that this trip was done by bicycle and more likely that the application was running and collecting
data while the trip was being done by motor-vehicle since the speed is far too high. All trips with an average speed greater than 30km/h are dropped from the analysis.

**Slow speed:** Using the average speed for each entire trip, if the average speed is less than 1 km/h, it is very likely that the application was left running for hours after the trip ended and the cyclist reached their destination. All trips with an average speed less than 1km/h are dropped from the analysis.

**Short trip:** Looking at the number of points making up the trip or the trip duration, if the trip lasted fewer than 60 seconds, this trip is too short for analysis. All trips shorter than one minute are dropped from the analysis.

**Time gap:** If a cyclist arrives at a signalized intersection as the signal changes to red, this cyclist would need to wait no more than 1 minute. To be conservative, we are assuming a value of 90 seconds meaning that if a trip stops for longer than 90 seconds, it is likely the cyclist stopped to run an errand and then continued riding but never stopped and re-started the application. In some cases, the time difference between points is several minutes or hours which may imply that the cyclist made a stop and went indoors (which explains why the GPS did not report any coordinates and timestamps during that entire time). In many cases, it seems some cyclists biked to work and left the application running for the entire 8-hours they were at work. All trips where the time between any two points in that trip is greater than 90 seconds are dropped from the analysis.

### 5.4.2. Separate Intersections and Segments

In this work, we look at intersections and segments separately. The geospatial file (shapefile) used for the intersections, defines intersections as the location (point) where the line segments intersect. The line segment shapefile are the lines down the centreline of each road. In order to capture the portions of the road which are associated with the intersection, a buffer distance of 20 metres is chosen. This distance is the same as is used to associate accidents to intersections. This distance can be constant or a variable depending on the size of the intersection. Different buffer dimensions which vary according to the size of each intersection will be developed in future work. In this study, a circular buffer of 20 metres is therefore defined around each intersection defining its area and the road segments falling outside of these buffers are considered as the segments. These buffers are shown in Figure 5-2. Square buffers make more intuitive sense but are very complicated to construct in ArcMap. A test was done on a sample of the data to compare the results using both
circular and square buffers and the difficulty in constructing square buffers properly aligned and rotated to match each intersection was not found to be justifiable, while circular buffers are sufficient and easy. Also, using this buffer size, there are on average 7 GPS points per trip falling inside each intersection.

![Figure 5-2 Sample of the Road Network showing the Separation of Intersections and Segments](image)

5.4.3. **Compute Instantaneous Speeds**

The instantaneous speed, $v_{j,i}$, between two consecutive positions, $i$ and $i-1$ for trip $j$ is computed as shown in the following equation.

$$v_{j,i} = \frac{\sqrt{(x_{j,i} - x_{j,i-1})^2 + (y_{j,i} - y_{j,i-1})^2}}{t_{j,i} - t_{j,i-1}}$$

For each trip, the distance travelled between consecutive positions, with coordinates, $(x_{j,i}, y_{j,i})$ and $(x_{j,i-1}, y_{j,i-1})$, are computed. For most cases, the GPS recorded the location every second, but especially in downtown, due to the density of tall buildings, the GPS may have recorded in 2 second intervals. If the time between consecutive points is greater than 1 second, the
instantaneous speed is computed as the distance traveled between consecutive points divided by the time it took to travel from point $i-1$ to point $i$ represented by $t_{j,i} - t_{j,i-1}$. It should be noted that the data has already been cleaned of excessively long time intervals between consecutive points.

5.4.4. **Compute Instantaneous Accelerations and Decelerations**

Similar to computing the instantaneous speeds, the instantaneous accelerations and decelerations, represented by $D_{j,i}$, can be computed as the difference in instantaneous speeds as shown in the following equation. A negative value for $D_{j,i}$ represents a deceleration between consecutive points.

$$D_{j,i} = \frac{v_{j,i} - v_{j,i-1}}{t_{j,i} - t_{j,i-1}}$$

According to the Federal Highway Administration, the 85th percentile of cyclist reaction time to stop is 1.3 seconds. One second is added to this value and is then rounded to 2.5 to be conservative (Landis et al., 2004). Using this value and computing the 85th percentile of cyclist deceleration rate is 3.3 m/s$^2$. Based on this value, any deceleration rate greater than 3.3 m/s$^2$ is defined as being dangerous and may be a sign of conflict. Another report shows this value to be 3.4 m/s$^2$ so this is the value used in this study (Pein, 2004). A sample of the data points along segments and through intersections is shown in Figure 5-2, identifying points with dangerous decelerations represented by red points along segments and pink points through intersections. Figure 5-3 shows a sample of acceleration/deceleration data for a randomly selected portion of a trip from second 1370 of that trip to 90 seconds later. In Figure 5-3, the dotted red line represents the dangerous threshold. Points falling below this line are defined as being dangerous. Dangerous decelerations as well as the EB of injuries can be used to identify hotspots and to rank the sites.
5.4.5. Empirical Bayes (EB) Estimator

The simplest way to compare decelerations to injuries is by using raw injury data as it has been recorded by the police or from ambulance interventions. An even better way is to compute the expected number of injuries using statistical models and then computing the empirical Bayes (EB) estimator, this way combining both the observed and predicted injuries which is an improvement from using each piece of information separately. The EB estimator, $EB_{\theta}$, is defined as:

$$EB_{\theta} = E(\theta_i|y_i) = (1 - w_i)y_i + w_i\mu_i \quad \text{with} \quad w_i = \frac{\phi}{\phi + \mu_i}$$

Where subscript $i$ represents the signalized intersection, non-signalized intersection or segment, $y_i$ represents the number of observed injuries, $\mu_i$ represents the number of predicted injuries from the model, $w_i$ represents the weighing factor which puts more weight on the observed or predicted injuries depending on the dispersion parameter from the model, $\phi$.

From our previous work, we developed safety performance functions for all signalized and non-signalized intersections as well as for road segments. (Strauss et al., 2015). Applying these developed models, we are able to compute the EB estimator which will compared to the decelerations.
5.4.6. Explore the Relationship between Injuries and Deceleration Rate

Ideally, we would like to be able to identify dangerous sites or hotspots prior to waiting several years to witness a large number of injuries. Using the Spearman rank correlation coefficient (also known as Spearman’s rho) we can compare the ranking of sites based on observed injuries and based on deceleration rate in order to explore the use of deceleration rate as a surrogate safety measure. Spearman rank correlation coefficient, \( \rho \), is defined as the Pearson correlation between ranks and is computed as shown in the following equation.

\[
\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}
\]

Where, \( d_i = Inj_i - DR_i \) which is the difference between the ranks of site \( i \) based on injuries, \( Inj_i \), and deceleration rate, \( DR_i \). \( n \) is the number of observations or sites which in this case is the number of intersections or segments.

5.5. SITE SELECTION AND DATA

The data sources are GPS traces and historical injury data.

5.5.1. Smartphone GPS Trips and Traces

This work uses GPS cyclist trip data collected from the Mon RésoVélo Smartphone application for both android and iOS Smartphones (Jackson et al., 2014). When cyclists begin their trip, they start the application which records their second-by-second latitudes, longitudes and timestamps (depending on the phone and the quality of the GPS signal, these may be recorded less often). This work uses data collected from July 2\textsuperscript{nd}, 2013, which is the day the application was launched, until November 15\textsuperscript{th}, 2013, which represents the day when most of the bicycle facilities in Montreal close for the winter. The study period is therefore 137 days long and in this short time, over 10,000 trips were recorded by almost 1,000 cyclists, representing over 16 million GPS data points. For the purpose of this study, all intersections and road segments without any GPS trips during the study period have been dropped. Combining GPS trip data with manual and automatic bicycle count data previously available, average annual daily bicycle flows (AADB) at each intersection and segment have been obtained (Strauss et al., 2015).
5.5.2. Historical Cyclist Injury Data

This study focuses on cyclist injuries which occurred at intersections and along road segments over a six-year study period from 2007 to 2012 obtained from police reports. This is the most recent accident data that is available and has been geo-coded for mapping. Table 5-1 shows the distribution of raw cyclist injury data at intersections and segments on the island of Montreal. Accidents are considered as having occurred at an intersection if they are within 20 metres of the centre point of the intersection and are otherwise assigned to road segments. During this time period, over 4,000 cyclists were injured at intersections and over 900 were injured along segments. Intersections and segments in the central areas of Montreal witnessed more injuries. This study also uses the EB expected number of cyclist injuries and risk obtained from the authors’ previous work (Strauss et al., 2015). The benefit of using expected EB injuries over raw injury data is that the expected injuries adjust the random events based on bicycle and motor-vehicle flows (as a proxy) as well as geometric design characteristics.

<table>
<thead>
<tr>
<th>Number of accidents</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signalized intersections</td>
</tr>
<tr>
<td>0</td>
<td>59.67</td>
</tr>
<tr>
<td>1</td>
<td>22.15</td>
</tr>
<tr>
<td>2</td>
<td>8.22</td>
</tr>
<tr>
<td>3</td>
<td>4.26</td>
</tr>
<tr>
<td>4</td>
<td>2.32</td>
</tr>
<tr>
<td>5</td>
<td>1.26</td>
</tr>
<tr>
<td>6</td>
<td>0.87</td>
</tr>
<tr>
<td>7</td>
<td>0.39</td>
</tr>
<tr>
<td>8</td>
<td>0.24</td>
</tr>
<tr>
<td>9</td>
<td>0.10</td>
</tr>
<tr>
<td>10</td>
<td>0.19</td>
</tr>
<tr>
<td>11</td>
<td>0.10</td>
</tr>
<tr>
<td>12</td>
<td>0.05</td>
</tr>
<tr>
<td>13</td>
<td>0.05</td>
</tr>
<tr>
<td>14</td>
<td>0.10</td>
</tr>
<tr>
<td>25</td>
<td>0.00</td>
</tr>
<tr>
<td>26</td>
<td>0.05</td>
</tr>
</tbody>
</table>
5.6. RESULTS

The average speed and acceleration/deceleration have been obtained for all intersections and segments on the island of Montreal. A summary of this information is provided in Table 5-2. Figure 5-4 shows the speed distribution for segments and intersections. Once these were computed, the dangerous decelerations could be identified. The results are discussed in the following sections.

Table 5-2 Summary of Speed and Acceleration/Deceleration for Intersections and Segments

<table>
<thead>
<tr>
<th></th>
<th>Intersections</th>
<th>Segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Speed (Km/h)</td>
<td>17.6</td>
<td>19.1</td>
</tr>
<tr>
<td>Average Acceleration/Deceleration (m/s²)</td>
<td>0.07</td>
<td>-0.04</td>
</tr>
<tr>
<td>Average Number of GPS points</td>
<td>46</td>
<td>41</td>
</tr>
</tbody>
</table>

5.6.1. Intersections

The average number of cyclist injuries over the 6 year study period using the EB method is 0.83 at signalized intersections and is 4 times smaller (0.22) at non-signalized intersections. 72% of non-signalized intersections have an EB value below the mean value. There are many more non-signalized intersections than signalized ones and far fewer injuries at non-signalized intersections. For these reasons, the correlation of the ranks was carried out separately for both types of intersections to best explore the correlation between injuries and decelerations for each type of facility.

From Figure 5-5 we can see that there exists a trend between the expected number of EB injuries and the number of dangerous decelerations whereas the trend between the ratio of
dangerous decelerations to the number of trips and EB risk do not seem to follow a similar trend. In terms of both dangerous decelerations and EB injuries, the maps show that these are both higher in the central areas of Montreal. The map of dangerous decelerations also highlights the streets which are sloped. Montreal in general has flat terrain except for around Mount Royal in the city’s centre. It is possible that streets that are sloped are in fact more dangerous since visibility is likely to be obstructed or not ideal or sufficient due to the slope. Figure 5-5 also shows heat maps for the number of dangerous decelerations as a rate per trip and EB injuries as a rate per AADB. Both maps show that in terms of rates, accounting for bicycle traffic, rates are greater outside the central areas.

Based on Spearman’s correlation, the ranks of EB injuries and dangerous decelerations have a correlation of 0.55 at signalized intersections and 0.37 at non-signalized intersections. By isolating only the non-signalized intersections in the central business district (CBD) of the island (3,526 intersections, representing 18% of the total), the correlation reaches a value of 0.55. For signalized intersections, 25% are located in the CBD and the correlation does not change.

The mean number of dangerous decelerations at signalized intersections is over 2.5 times greater than at non-signalized intersections, as shown in Table 5-3. This also confirms that signalized intersections are more dangerous than non-signalized intersections, in terms of both observed injuries and deceleration rate as a surrogate safety measure.

| Table 5-3 Summary Statistics for Signalized and Non-Signalized Intersections |
|-------------------------------------------------|----------------------------------|----------------|----------------|----------------|
| **Type**                    | **Variable**                  | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| Signalized                  | Number of trips               | 52.3     | 73.5             | 1            | 487          |
|                             | AADB                           | 1381     | 3139            | 8.4          | 28961        |
|                             | Dangerous deceleration        | 12.65    | 18.68           | 0            | 154          |
|                             | Number of injuries            | 0.79     | 1.44            | 0            | 14           |
|                             | EB injuries                   | 0.76     | 0.89            | 0.08         | 19.04        |
|                             | Number of sites               | 2023     |                 |              |              |
| Non-Signalized              | Number of trips               | 21.3     | 43.4            | 1            | 574          |
|                             | AADB                           | 311      | 1047            | 17.4         | 46393        |
|                             | Dangerous deceleration        | 4.92     | 10.09           | 0            | 149          |
|                             | Number of accidents           | 0.19     | 0.64            | 0            | 13           |
|                             | EB injuries                   | 0.22     | 0.36            | 0.012        | 8.52         |
|                             | Number of sites               |          |                 | 11256        |              |
5.6.2. Road Segments

The average number of injuries using the EB method is 0.066 which is over 12 times lower than for signalized intersection and over 3 times lower than for non-signalized intersections. Over 78% of segments have an EB value below the mean value. Summary statistics for road segments are shown in Table 5-4.

Again one can see, in Figure 5-6, the similarity in the heat maps for segments based on dangerous decelerations and EB injuries. Also, even though the number of dangerous decelerations and EB injuries are lower at segments than at intersections, Figure 5-6 still shows a similar trend that these are both concentrated in the CBD of Montreal. For segments, even in terms of injury rate using the EB method, these also appear to be greater in the CBD and more spread out for deceleration as a rate per trip.
Based on Spearman’s correlation EB injuries and dangerous decelerations their ranks have a correlation of 0.48 for road segments. By isolating the segments in the CBD only (1,249 intersections, representing 11% of the total), the Spearman correlation reaches a value of 0.58.

**Table 5-4 Summary Statistics for Road Segments**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of trips</td>
<td>22.3</td>
<td>48.7</td>
<td>0</td>
<td>713</td>
</tr>
<tr>
<td>AADB</td>
<td>1851</td>
<td>676</td>
<td>1387</td>
<td>16016</td>
</tr>
<tr>
<td>Dangerous deceleration</td>
<td>16.34</td>
<td>48.69</td>
<td>0</td>
<td>1375.0</td>
</tr>
<tr>
<td>Number of accidents</td>
<td>0.03</td>
<td>0.38</td>
<td>0</td>
<td>25.0</td>
</tr>
<tr>
<td>EB injuries</td>
<td>0.066</td>
<td>0.09</td>
<td>0.02</td>
<td>2.56</td>
</tr>
<tr>
<td>Number of sites</td>
<td></td>
<td></td>
<td>19837</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5-6 Heat Maps of Dangerous Decelerations, EB Injuries and Rates on Segments**
Figure 5-7 Number of EB Injuries per Km (top) and Number of Dangerous Decelerations per Km (bottom) for Corridors
5.6.3. Corridors

In addition to looking at individual road segments, the segments can be aggregated into corridors. All the roads classified as being arterials, have been aggregated into corridors as shown in Figure 5-7. The top 10 most dangerous corridors based on both the number of EB injuries per kilometre and the number of dangerous decelerations per kilometre were then compared. It was found the 6 of the 10 top corridors overlapped for the ranking based on both criteria.

5.7. CONCLUSION AND FUTURE WORK

In this work, we investigate the use of deceleration rate to detect evasive actions taken in order to avoid a collision, as a surrogate safety measure for cyclists, obtained from GPS data. That is, this study developed a methodology to use cyclist GPS trip data from a Smartphone application to extract hard decelerations in the entire network of roads and intersections. Decelerations are then correlated with injuries estimated using the EB approach. The relationship between decelerations and accidents demonstrates its potential use as a surrogate safety measure. The benefits of using surrogate safety measures as opposed to historical accident data are several, most importantly, analysis and implementation of treatments based on accident data represent a reactive approach to road safety, whereas surrogate measures enable a proactive approach to identify dangerous locations prior to witnessing a significant number of accidents which can cause grievous bodily harm. There is therefore a need to define events that occur more frequently than accidents, are observable and can be related to accidents and safety.

According to Spearman’s rank correlation coefficient and based on the heat maps, we can see that the number of dangerous decelerations are significantly correlated with the number of injuries from the EB method based on the safety performance functions developed for signalized intersections, non-signalized intersections and segments using historical accident data and accounting for bicycle flow, geometric design and proxies for motor-vehicle flow. The correlation coefficients for signalized intersections (value of 0.55) are slightly higher than for segments (value of 0.48). Also, the number of cyclist injuries and dangerous decelerations are much greater in the CBD for non-signalized intersections and segments. For non-signalized intersections, the correlation coefficient increases from a value of 0.37 for the entire island of Montreal to 0.55 when the CBD is isolated. For segments, the value increases from 0.48 to 0.58.
These results show that deceleration rate, obtained from GPS data, can be used as a surrogate safety measure for cyclists, to identify sites that are dangerous prior to witnessing a high number of injuries at these sites. Using the Smartphone application GPS data, deceleration rates can be used to identify hotspots for road segments and intersections in the entire network and identify locations needing further inspection and potentially safety treatments.

As part of future work, we will carry out a sensitivity analysis and try different thresholds for identifying dangerous decelerations and the slopes of the road segments will be accounted for. Also different buffer dimensions for separating segments and intersections will be tested. The Smartphone application is still in use so this work can be updated using recent cyclist data and it is likely that some roads and intersections which did not witness any trips during the study period presented in this paper will now have trip data. Also cyclist behaviour is an important factor to consider for surrogate safety measures such as deceleration rate. For aggressive cyclists, accelerating and decelerating hard may be normal and not necessarily related to conflict involvement. Using this same GPS data we can account for cyclist behaviour provided that each cyclist recorded some minimum number of trips using the application.

5.8. AKNOWLEDGEMENTS

We acknowledge the financial support provided by the FRQNT and NSERC. We would like to thank the Montreal Department of Public Health and Urgences-santé for collecting and validating the injury data as well as Brisk Synergies and the city of Montreal for the GPS data. All remaining errors and the views expressed in this research are, however, solely ours.

5.9. REFERENCES


Ville de Montréal, n.d. Mon RésoVélo [WWW Document]. URL
http://ville.montreal.qc.ca/portal/page?_pageid=8957,112451619&_dad=portal&_schema=PORTAL


Link between Chapter 5 and Chapter 6

All the previous chapters in this thesis focused on safety. Safety analysis was carried out at a sample of sites for cyclists, pedestrians and motor-vehicles and safety analysis was carried out at the entire population of segments and intersections, based on observed injuries and deceleration rate as a surrogate safety measure using the Smartphone application data.

In addition to safety applications, the Smartphone application can also be used for cyclist comfort and level-of-service applications. Using this source of data, cyclist speeds along segments as well as delays through intersections can be extracted. This knowledge serves as vital information for navigation and routing purposes since travel times and speeds can be measured at all times of the day and for all trip purposes to provide accurate travel times for cyclists from origin to destination instead of assuming a constant travel speed throughout the day and for all roads.
Chapter 6

Speed, Travel Time and Delay for Intersections and Road Segments in Montreal using Cyclist Smartphone GPS Data
Chapter 6: Speed, Travel Time and Delay for Intersections and Road Segments in Montreal using Cyclist Smartphone GPS Data

Jillian Strauss\textsuperscript{a} and Luis Miranda-Moreno\textsuperscript{a}

\textsuperscript{a}Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, Canada

6.1. ABSTRACT

Until now, very little is known about cyclist speeds and delays at the disaggregate level of each road segment and intersection. Speed and delay serve as vital information for navigation and routing purposes since they can identify speeds and delays during different times of the day and how they differ across roads and bicycle facilities. In this work, we explore the use of recent GPS cyclist trip data, from the Mon R\textsuperscript{e}soV\textsuperscript{e}lo smartphone application, for identifying different level-of-service measures such as travel time, speed and delay at the level of the entire road and intersection network for the island of Montreal. Also, a linear regression model is formulated to identify the geometric design and built environment characteristics affecting cyclist speeds on segments. Among other results, on average, segment speeds are greater along arterials than on local streets and greater along segments with bicycle infrastructure than those without. Modeling cyclist speed revealed that the variable representing the cyclists’ average speed on uphill, downhill and level segments, cyclists’ average speed on arterials as well as geometric design and built environment affect segment speeds. The model results identify that segments which have cyclists biking for work or school related purposes, segments used during morning peak, segments with bicycle infrastructure and segments which do not have signalized intersections at either end, tend to have cyclists riding at greater speeds. Also, cyclists travel faster when the temperature is between 10\textdegree{} and 20\textdegree{} Celsius and travel slower late at night or in the early morning.
6.2. INTRODUCTION

Cities have begun to see a rise in cyclist numbers and in response, have started to build and provide bicycle infrastructure. In Montreal, the bicycle facility network has almost doubled over the 10 years from 2000 to 2010 (Vélo Québec, 2013). In Montreal, over 700,000 people, representing 52% of the adult population, ride bicycles and in 2010, the modal share of cycling to work reached 2.2%. In the central neighbourhoods of the island during the summer months, the modal split of cycling reaches as high as 10%. Currently in Montreal, some cycle tracks are so widely used that they have begun to become congested. With growing issues of climate change as well as energy and health concerns, we have begun to see a shift away from motor-vehicle dominance and a rise in active modes of transportation such as cycling. In response to the increasing number of cyclists, cities are changing their approaches to infrastructure and signal phasing to apply a multimodal perspective. In order to propose appropriate design changes and signal phasing modifications to account for cyclists and not only vehicles, reliable travel time, speed, delay and other level-of-service measures are necessary for cyclists. Variations in these measures throughout the day is also an important component in accommodating all modes.

While many studies in the literature look at vehicle speeds, travel times, delays and level-of-service (LOS), very little is known about cyclist travel times and speeds along roads and bicycle facilities, and through intersections at the disaggregate level. Previous works attempting to tackle these measures, have been restricted by small samples of cyclists and study locations to draw meaningful conclusions. For the most part, these previous studies have focused specifically on arterials and have not tackled the complexity of intersections, other road types or bicycle facilities. Many studies have been based on surveys or GPS units for which both can affect cyclist behaviour and therefore affect these studies’ results and conclusions about the LOS measures. These measures serve as important information for cities regarding the quality of the accommodations provided for cyclists and can identify whether or not they are sufficient. These measures also have many applications in navigation and routing for cyclists which, until now assumes one value for speed for the entire network regardless of the type of facility, time of day, flow conditions or trip purpose.

This research aims to explore the use of recent cyclist GPS trip data, coming from the Mon RésoVélo smartphone application, to compute and evaluate different measures of cyclist level-of-service or “bikeability” on the island of Montreal and validate the reliability and usefulness of this source of data. In other words, we are interested in quantifying cyclist travel times, speeds and
delays along road segments, cycle tracks and through intersections and how these may vary by trip purpose and for different periods of the day. Due to the richness of the data, this work can be carried out at the entire network level and from a completely disaggregated approach, focusing on individual segments and intersections as the units of analysis. Also, unlike surveys and GPS units, this data is completely anonymous, so that cyclist behaviour is unlikely to be affected by the tracking of their trips.

This work has two main objectives, to: 1) develop a methodology to estimate cyclist speeds, travel times and delays in a road network using GPS data, and 2) identify the factors affecting these level-of-service parameters.

6.3. LITERATURE REVIEW

Cyclist delay, travel time, speed and level-of-service (LOS) are all vital components in determining how to best provide for cyclists along both road segments and at intersections. To date, studies addressing these components, have been carried out focusing on measures of cyclist LOS along arterial segments (Epperson, 1994; Harkey et al., 1998; Landis, 1994; Landis et al., 1997; Sorton and Walsh, 1994). Very few studies, if any, have addressed the complexity of urban intersections to determine these measures for cyclists. In order to be able to assess the quality of services for cyclists, measures of delay and LOS are required for both segments and intersections since both are road network elements that cyclists need to traverse to get from their origin to their destination. The Florida DOT set out to develop a bicycle LOS measure for through movements at signalized intersections (Landis et al., 1828). This was achieved by having participants cycle along a specified course in Orlando, Florida and providing their opinions about each intersection they passed through. Using this data, the Florida DOT sought to develop a mathematical model to express cyclists’ LOS and level of accommodation through intersections. A model having an R-squared value of 0.83 was developed which identified that the width of the outside through lane has a negative effect on LOS (improves LOS) whereas crossing distance and vehicle volume have a positive effect on LOS (worsen LOS). One study in Albany, California, aimed to estimate intersection delays using a virtual trip line (VTL) method based on cell phone GPS through one intersection (Ban et al., 2009). VTLs were installed both upstream and downstream of each intersection to obtain travel times between both points. Another study carried out in Minneapolis, Minnesota, studied the travel speeds of a small sample of cyclists riding along different types of
bicycle facilities (El-Geneidy et al., 2007) from GPS data. Travel speeds can be used as an input in determining accessibility measures for cyclists.

While these previous studies are useful in identifying the factors affecting cyclist LOS, these were identified for one intersection or a small sample of intersections and segments. Overall, studies have been carried out to obtain travel times using new data collection methods, such as Bluetooth (Mei et al., 2012). Also the impact of travel time on cyclist route choice has also been investigated through stated preference surveys (Hunt and Abraham, 2007; Sener et al., 2009; Stinson and Bhat, 2003). However, other sources, such as GPS data, have not been widely tested.

A new data collection method that has begun to gain popularity in the research is GPS based smartphone applications. These applications are still very new and few studies have been carried out looking into cyclist travel times, speeds, delays and LOS. GPS data has been used previously but by equipping a small sample of cyclists with GPS equipment. One study in Oregon, Portland, collected GPS data from 164 cyclists for 8 months of the year (Dill and Gliebe, 2008). Among other results, cyclists riding for utilitarian purposes were found to ride mainly on bicycle infrastructure and cyclists generally do not take the shortest path and deviate to ride on bicycle infrastructure and low-traffic streets.

Overall, previous studies have relied on a small sample of cyclists. Some studies have applied GPS data to model cyclist route choices and preferences as well as speeds and travel times and some have even combined the GPS data with surveys to obtain cyclists’ opinions about specific routes. One benefit of GPS data coming from smartphone applications is that they have the potential to provide route data for all cyclists possessing a smartphone, providing the potential for a much larger sample of cyclists and a much wider variety of routes to study and from which to extract travel times, speeds and delays. Also, the application is more discrete than equipping some cyclists with GPS units. With GPS units, the cyclists agreed to be part of a study with a specific goal and therefore this may influence their behaviour. Whereas with the smartphone application, cyclists are logging their trips and benefit from obtaining a view of the route they have taken including their distance travelled, average and maximum speeds, greenhouse gas emission savings and energy expended through a calorie counter. All this data is collected and remains anonymous.

In 2010, the San Francisco Municipal Transportation Agency (SFMTA) launched, CycleTracks, a smartphone application to collect cyclist trip data to gauge their current bicycle infrastructure and help guide future planning. The SFMTA used this application to develop travel
demand models to study where cyclists are riding and how bicycle infrastructure affects their route choices. Some results have been published and identify that cyclists prefer bicycle lanes over other facility types and are discouraged by steep slopes, turning and long distances (Hood et al., 2011). Following San Francisco’s success, CycleTracks has been implemented in 10 cities in the United States. Also, about 7 cities, including Montreal and Atlanta, have modified this open source application to meet their city’s needs.

Atlanta launched Cycle Atlanta in hopes of collecting data to aid in improving cycling conditions and developing a connected network of bicycle facilities. This application also enables cyclists to comment on trouble areas along their trip.

This study is the first of its kind to apply GPS data from a smartphone application to study travel times, speeds and LOS. This GPS data for the island of Montreal has never been analyzed before. Also, these measures have previously been studied only for vehicles. This work is one of the first to study these measures for cyclists and at the disaggregate level of segments and intersections for an entire urban road network.

6.4. METHODOLOGY

The methodology has several steps, following the flowchart in Figure 6-1.

Steps:
1) Compute distances, times and speeds from raw GPS data
2) Develop and apply data cleaning rules
3) Apply a speed smoothing technique
4) Separate segments and intersections
5) Average speed on segments and speed model
6) Obtain travel time through intersections
7) Compute intersection delay
8) Develop a segment speed model
The following steps are carried out for all trips, riding along the corridors and through the intersections in the boroughs designated as downtown Montreal.

**Figure 6-1 Methodology Flowchart**

The following steps are carried out for all trips, riding along the corridors and through the intersections in the boroughs designated as downtown Montreal.
6.4.1. Distance, Time and Speed Computations

Each trip in the dataset is made up of second-by-second raw GPS data. Consider trip \( t \), which is made up of GPS points, \( p = (p_1, p_2, \ldots, p_k) \), with \( k \) representing the number of points making up trip \( t \), and the \( i^{th} \) point is defined by \( p_i = (x_i, y_i, r_i) \) where \( x_i \) is the latitude, \( y_i \) is the longitude and \( r_i \) is the timestamp.

The first step involves converting the point coordinates from a World Geodetic System with locations in latitude and longitude to a Projected Coordinate System with XY coordinates in metres. In other words, changing the projection to NAD 1983 Zone 8, adjusts for the curvature of the Earth to map the points in two dimensions. With the coordinates in metric units, the Euclidean distance, \( d_{ij} \), in metres, between consecutive GPS points can be computed as:

\[
d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}
\]

Where, \( i \) and \( j \) are consecutive points and the cyclist is traveling from \( i \) with \( p_i = (x_i, y_i, r_i) \) to \( j \) with \( p_j = (x_j, y_j, r_j) \).

Using the timestamps, the time difference, in seconds, between each pair of consecutive points can be computed as:

\[
w_{ij} = (3600 \times h_j + 60 \times m_j + s_j) - (3600 \times h_i + 60 \times m_i + s_i)
\]

Where, \( h \) represents the hour, \( m \) represents the month and \( s \) represents the second.

For distance and time difference computations, the first point of each trip is assigned a value of zero for both distance and time and the values of distance and time difference are computed from that point onward. Combining the computed distances and times, the speed, \( v_{ij} \), in meters per second (m/s) can easy be calculated and then converted into units of kilometres per hour (km/h).

\[
v_{ij} = \frac{d_{ij}}{w_{ij}} = \frac{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}}{(3600 \times h_j + 60 \times m_j + s_j) - (3600 \times h_i + 60 \times m_i + s_i)}
\]

6.4.2. Data Filtering

Due to the noise of the GPS data as a result of buildings (especially in downtown Montreal where there are many high-rise buildings), trees and adverse weather as well as the accuracy of the
Smartphone’s GPS and if it malfunctions, the data needs to be checked and cleaned of the noise. After thorough inspection of the data, the following cleaning rules have been set up:

1. **High speed**: Using the instantaneous point-to-point speeds, $v_{ij}$, already computed, the average speed for each entire trip $t$, $v_t$, is computed as:

   $v_t = \frac{\sum_{i,j} v_{ij}}{k_t}$

   Where, $k_t$ is the total number of points in trip $t$. Then if $v_t > 30\text{km/h}$, the trip is classified as a non-bike trip since it is very unlikely that this trip was in fact done by bicycle and more likely that the application was running and collecting data while the trip was being done by motor-vehicle. Then a simple rule is set up as: drop if $v_t > 30\text{km/h}$.

2. **Slow speed**: In a similar way to the high speed filtering rule, trips with $v_t < 1\text{km/h}$ are excluded since it is very likely that the application was left running for hours after the trip ended and the cyclist reached their destination. A simple rule is set up as: drop if $v_t < 1\text{km/h}$.

3. **Short trip**: Looking at the number of points, $k$, making up the trip or based on the sum of the instantaneous time differences for each trip, if the trip lasted fewer than 60 seconds, this trip is too short for our interests. A simple rule is set up as: drop the entire trip if the total length of the trip, in time, is less than 1 minute.

4. **Time gap**: If a cyclist arrives at a signalized intersection as the signal changes to red, at most this cyclist would need to wait no more than 1 minute. To be conservative, we are assuming a value of 90 seconds meaning that if a trip stops for longer than 90 seconds, it is likely the cyclist stopped to run an errand and then continued riding but never having stopped and re-started the application. In some cases, the time different between points is several minutes or hours which may imply that the cyclist made a stop and went indoors (which explains why the GPS did not report any coordinates and timestamps during that entire time). In many cases, it seems some cyclists biked to work and left the application running for the entire 8-hours they were at work. Trips that stop for a long period of time and then start up again can be split into two separate trips but additional cleaning may be needed once treating the trip as two separate trips. To avoid this additional work to develop our initial methodology, we have set up a simple rule: drop if $w_{ij} > 90$ seconds.
6.4.3. Speed Smoothing

Speed along segments is an important and necessary input to compute the delay through intersections, a smoothing method is therefore applied to obtain reliable speeds along segments for the bike trips that remain following the cleaning rules. The raw GPS data has many peaks and troughs due to the accuracy of the Smartphones GPS therefore a simple filtering technique is used which is based on the median of some speed values before and some speed values after the current instantaneous speed. In this work, we tested three filters: median of 3 values, median of 5 values as well as the median of 7 values, represented as:

\[ v_m = \left( \frac{n + 1}{2} \right)^{th \ term} \]

Where \( v_m \) is the median speed and where \( n \) can have three different values: 3, 5 and 7. In order to use the above equation, if applying the median of 3 values filter, the speed values of one point before the current point, one point after the current point and the speed of the current point first need to be sorted. Once sorted, the equation identifies that the second value (after sorting) is the value taken and assigned to the current position. Similarly, for the median of 5 values filter, the speed values from two points before, from two points after, as well as the current speed value are sorted and the third speed is taken as the current speed. Again, for the median speed of 7 values filter, the speed values of three points before the current speed, three points after the current speed and the current speed are sorted and the fourth speed value is used as the new current speed.

In Figure 6-2, a portion of a trip was selected to show the unfiltered and filtered speeds, showing the peaks and troughs resulting from the noisy data and how the median filters smooth them. The smoothing effects of the different filters can be compared graphically. Looking at this graph we can see how well this simple filtering method smoothens out the peaks in the speed data. The filtered and unfiltered speeds will later be tested for sensitivity.
6.4.4. Separate Segments and Intersections

Now that we have cleaned the data and smoothed out the instantaneous speeds, we want to separate the portions of each trip done along segments and within the vicinity of intersections. A segment is defined as the link between neighbouring intersections and each has its own unique ID. An intersection is defined as the point at the intersection of these links and also has its own unique ID. Segments and intersections are separated by generating 20 metre buffers around all intersections and road segments separately. In order to have completely separate buffers for intersections and segments, the intersection buffers need to be removed from the segment buffers where they overlap. 20 metre buffers were selected around intersections to include the potential effect that the traffic signal has on cyclist speed at 20 metres away from the intersection. In other words, within this buffer dimension, the cyclist’s reaction and behaviour in the presence of the traffic signal, including deceleration, stopping (if the cyclist stopped) and acceleration are all associated with the intersections and do not affect the segment speeds.

The trip points are then intersected with the intersection and segment buffers and each point is assigned to a specific intersection or road segment. There are now two datasets, one with all the trip points along road segments and the other one with all trip points in the vicinity of intersections. From this point forward, separate analysis is carried out for both datasets.
6.4.5. Average Speed on Segments

The methodology developed to obtain delays through intersections requires first that the travel speeds along segments be known. By identifying the first and last points of each trip along each segment, the travel direction can be extracted so that the speed will be computed separately for both directions. The instantaneous speed values for all points per trip, per segment and per direction are averaged to obtain average speeds. The speed along each segment-direction, \( v_s \), is computed as:

\[
v_s = \frac{\sum v_{d,t} n_d}{n_d}
\]

Where, \( v_{d,t} \) represents the segment speeds for each trip and \( n_d \) represents the total number of trips that travelled along that road segment and in the same direction.

To obtain the speeds per segment and per direction, the speeds for all trips using each segment and traveling in the same direction are averaged. These calculations are carried out four times, once for the unfiltered speed and for the median of 3, median of 5 and median of 7 filtered speeds.

Further cleaning was carried out at this step. Segment-directions with fewer than 2 trips passing through is not enough data to provide an accurate average speed. All segment-directions with fewer than 2 trips are dropped.

6.4.6. Travel Time through Intersections

The dataset of points within the vicinity of intersections includes both signalized and non-signalized intersections. For the purpose of finding delays, only signalized intersections are of interest. All non-signalized intersection data is therefore dropped from the intersection dataset. As an initial analysis, only the delays for through moving cyclists are considered. Once the methodology is developed, this work will be expanded to include all boroughs of the island of Montreal and all cyclist movements through intersections (left and right turns).

Using the timestamp of the GPS data and sorting the data in chronological order, the first and last points of each trip in each intersection can be identified in order to determine the direction of travel.
At this step, additional data cleaning is also carried out. The following cleaning rules are applied:

1. **Few points**: If there are fewer than 3 points per trip inside a given intersection buffer, this trip data is dropped for this intersection only, not the entire trip.

2. **Short distance**: The diameter of the buffer is 40 metres so the distance travelled per trip per intersection buffer should be as close as possible to 40 metres. Due to the noise of the GPS data, the distances travelled vary. Trip data is dropped if the total distance between the first and last points inside the intersection is less than 20 metres, implying that the cyclist passed through the intersection along the outer regions of the buffer and there are likely not many points falling inside the buffer.

3. **Long time**: The trip in the intersection is also dropped if the total time spent in the intersection is greater than 3 minutes. Worse case, if a cyclist arrives as the signal turns red, the signal should not remain red for more than 90 seconds. However, if the cyclist is stuck behind a queue of vehicles, the cyclist’s waiting time can be slightly longer in the worst case. This step intends to remove the stops that were not made as a result of the traffic signal but more likely to run an errand.

6.4.7. Intersection Delay

The segment speed data and the intersection travel time data per direction need to be matched in order to compute the intersection delays. The segment speed in the direction heading towards the intersection represents the theoretical speed that the cyclist would pass through that intersection if there was in fact no intersection there at all.

Delay, $D_i$, is computed in terms of time in seconds based on Equation 1. In other words, delay is the time difference between traveling through the intersection buffer at the segment speed and the intersection speed. The resulting time difference represents the delay introduced to the cyclist’s trip as a result of the traffic signal per intersection and direction of travel.

$$D_i = \left( \frac{c_m}{V_{int}} - \frac{c_m}{V_{seg}} \right) \times 3.6 \quad (1)$$

Where, $V_{int}$ is the average speed through the intersection in km/h, $V_{seg}$ is the average speed along the segment in km/h, $c_m$ is the diameter of the intersection buffer, which is 40 metres in this
study (but other values for \( c_m \) can be used) and 3.6 represents the factor needed to convert the speeds from km/h to metres per second (m/s). Also, the delay in Equation 1 is computed four times, based on the unfiltered and filtered speeds (median of 3, median of 5 and median of 7) to test the sensitivity of the results.

This paper serves to develop a methodology which can be applied to any city or road network, both large and small.

6.4.8. Segment Speed Model

To identify the geometric design and built environment characteristics affecting cyclist speeds on segments, a linear regression model is formulated as shown in Equation 2.

\[
y_i = \beta_i X_i + \epsilon_i
\]

Where \( y_i \) is the speed on the segment averaged across all users and trips which travelled on that segment during the study period, \( X_i \) represents flows, geometric design and built environment characteristics, weather conditions, trip and cyclist characteristics as well as cyclist behaviours and personality attributes and \( \beta_i \) is a vector of regression parameters estimated from the data. The error term, \( \epsilon_i \), represents an unobserved random variable.

6.5. SITE SELECTION AND DATA

This work focuses on the downtown boroughs of the island of Montreal: Plateau Mont-Royal, Ville Marie and Outremont, covering an area of over 27 square kilometres. These boroughs were selected to develop the methodology since there are many cyclist trips in this area.

6.5.1. Smartphone GPS Trips and Traces

This work uses GPS cyclist trip data collected from the Mon RésoVélo Smartphone application for both android and iOS Smartphones (Jackson et al., 2014). When cyclists begin their trip, they start this application which records their trips and provides second-by-second latitudes, longitudes and timestamps (depending on the phone and the quality of the GPS signal, these may be recorded less often). This work uses data collected from July 2\(^{nd}\), 2013, which is the day the application was launched, until November 15\(^{th}\), 2013, which represents the day when most of the bicycle facilities in Montreal close for the winter. The study period is therefore 137 days long and in this short time,
over 10,000 trips were recorded by almost 1,000 cyclists, representing over 16 million GPS data points needing to be processed to meet our objectives.

6.5.2. Weather Data

Weather data for Montreal, from the Pierre Elliott Trudeau Airport weather station, was downloaded from Environment Canada’s historical data available on their Website. The weather data downloaded is hourly temperature, relative humidity and wind speed data for the entire study period. The hourly weather data was matched to each trip based on the trip’s date and start time.

6.6. RESULTS

6.6.1. Segment Speed

The speeds along segments and through intersections were computed and their distributions are shown in Figure 6-3. After calculating the average speeds (unfiltered and filtered) per segment and per direction, these can be mapped. The filtered speed based on the median of 3 speed values is shown in Figure 6-4. In Figure 6-4, each segment is represented by two lines, one for each direction of travel. Segment-directions missing speed data appear as gray lines in Figure 6-4. Segment-directions with speed data range in colour from red to green, representing slow to fast speeds. Figure 6-4 also provides a zoom-in on a section of road to better see the speeds.

![Figure 6-3 Speed Distributions for Segments and Intersections](image-url)
Figure 6-4 Average Speed using Median 3 Filter on All Road Segments Downtown (top) and Zoom-In (bottom)
The average speeds (in km/h) by road type are provided in Table 6-1. The speeds in Table 6-1 make intuitive sense. The average speed on arterials is greater than the average speed on collectors which is in turn greater than the average speed on local streets. The average unfiltered speeds are in all cases greater than the filtered speeds. The GPS data is likely to be very noisy with many peaks in speed. This explains why the filtered speeds are lower since their purpose is to smooth out the peaks.

The average speeds (in km/h) by bicycle infrastructure type are provided in Table 6-1. The speeds in Table 6-1 also make intuitive sense. The average speed on cycle tracks is greater than the average speed on bicycle paths which is in turn greater than the average speed on segments without any bicycle infrastructure.

<table>
<thead>
<tr>
<th>Road Type</th>
<th>Unfiltered Speed</th>
<th>Filtered Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median of 3</td>
</tr>
<tr>
<td>Arterial</td>
<td>21.3</td>
<td>20.2</td>
</tr>
<tr>
<td>Collector</td>
<td>20.6</td>
<td>19.6</td>
</tr>
<tr>
<td>Local</td>
<td>19.5</td>
<td>18.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Road Type</th>
<th>Unfiltered Speed</th>
<th>Filtered Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median of 3</td>
</tr>
<tr>
<td>Track</td>
<td>21.6</td>
<td>20.6</td>
</tr>
<tr>
<td>Path</td>
<td>20.5</td>
<td>19.3</td>
</tr>
<tr>
<td>None</td>
<td>20.2</td>
<td>19.2</td>
</tr>
</tbody>
</table>

6.6.2. Intersection Delay

The developed methodology to compute delays for cyclists traveling through intersections is applied in downtown Montreal. A sample of the results is shown in Figure 6-5. In this figure the delays are shown for each intersection approach with sufficient data. The delay values are shown on the entering approach and range from green to red, representing short to long delays.

The delays computed for each approach can be averaged to obtain an average delay for the entire intersection. The average delays are also shown in Figure 6-5 again ranging from green to red, representing short to long delays.
Figure 6-5 Sample of Intersection Delays per Approach (top) and Averaged for all Intersections Downtown (bottom)
6.6.3. Segment Speed Model

Many variables were extracted or calculated from the data sources and their effects on segment speeds were tested. Table 6-2 provides a list of some of the variables which include bicycle flows, geometric design and built environment characteristics, weather conditions, trip and cyclist characteristics as well as cyclist behaviour and personality attributes.

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle Flow</td>
<td>Bicycle flow - segment</td>
<td>Bicycle flow (AADB) along the segment</td>
</tr>
<tr>
<td></td>
<td>Bicycle flow - intersection</td>
<td>Bicycle flow (AADB) at the intersection</td>
</tr>
<tr>
<td></td>
<td>a.m. bicycle flow</td>
<td>Bicycle flow (AADB) along the segment during morning peak</td>
</tr>
<tr>
<td></td>
<td>p.m. bicycle flow</td>
<td>Bicycle flow (AADB) along the segment during evening peak</td>
</tr>
<tr>
<td></td>
<td>Off peak bicycle flow</td>
<td>Bicycle flow (AADB) along the segment during off peak</td>
</tr>
<tr>
<td></td>
<td>Night bicycle flow</td>
<td>Bicycle flow (AADB) along the segment during the night</td>
</tr>
<tr>
<td>Geometric Design and Built Environment</td>
<td>Length of segment</td>
<td>In metres</td>
</tr>
<tr>
<td></td>
<td>Arterial or collector</td>
<td>Dummy variable = 1 if the road segment is classified as an arterial/collector</td>
</tr>
<tr>
<td></td>
<td>Bicycle facility</td>
<td>Dummy variable = 1 if the road segment has bicycle infrastructure</td>
</tr>
<tr>
<td></td>
<td>First signalized</td>
<td>Dummy variable = 1 if the intersection at the start is signalized</td>
</tr>
<tr>
<td></td>
<td>Last signalized</td>
<td>Dummy variable = 1 if the intersection at the end is signalized</td>
</tr>
<tr>
<td></td>
<td>Both signalized</td>
<td>Dummy variable = 1 if both intersections are signalized</td>
</tr>
<tr>
<td></td>
<td>None signalized</td>
<td>Dummy variable = 1 neither intersection is signalized</td>
</tr>
<tr>
<td></td>
<td>Uphill</td>
<td>Dummy variable = 1 if the slope is greater than +2% slope</td>
</tr>
<tr>
<td></td>
<td>Downhill</td>
<td>Dummy variable = 1 if the slope is less than -2% slope</td>
</tr>
<tr>
<td></td>
<td>Level</td>
<td>Dummy variable = 1 if the slope is between -2% and +2%</td>
</tr>
<tr>
<td>Trip Characteristics</td>
<td>Trip purpose</td>
<td>Dummy variable = 1 if work or school related</td>
</tr>
<tr>
<td></td>
<td>Weekday</td>
<td>Dummy variable = 1 if trip was made on a weekday</td>
</tr>
<tr>
<td></td>
<td>a.m. peak</td>
<td>Dummy variable = 1 if trip was made between 6 and 9 a.m.</td>
</tr>
<tr>
<td></td>
<td>p.m. peak</td>
<td>Dummy variable = 1 if trip was made between 3 and 7 p.m.</td>
</tr>
<tr>
<td></td>
<td>Off peak</td>
<td>Dummy variable = 1 if trip was made between 9 AM and 3 p.m.</td>
</tr>
<tr>
<td></td>
<td>Night</td>
<td>Dummy variable = 1 if trip was made between 7 PM and 6 a.m.</td>
</tr>
<tr>
<td>Weather Conditions</td>
<td>Temperature</td>
<td>In degrees Celsius - Different categories were selected, such as: 1) Ideal - between 18 and 25, 2) Cold - below 18 and 3) Hot - above 25</td>
</tr>
<tr>
<td></td>
<td>Relative humidity</td>
<td>Different categories were selected, such as: 1) Below 60, 2) Between 60 and 80 and 3) Above 80</td>
</tr>
<tr>
<td>Cyclist Characteristics</td>
<td>Age</td>
<td>7 categories of age: 1) Under 18, 2) 18-24, 3) 25-34, 4) 35-44, 5) 45-54, 6) 55-64 and 7) 65 and up</td>
</tr>
<tr>
<td></td>
<td>Male/female</td>
<td>Dummy variable = 1 if male and = 0 if female</td>
</tr>
<tr>
<td></td>
<td>Cycle daily</td>
<td>Dummy variable = 1 the cyclist bikes every day – measure of experience and familiarity with their route</td>
</tr>
<tr>
<td></td>
<td>Since childhood</td>
<td>Dummy variable = 1 if the cyclist has been cycling since childhood – measure of experience and comfort</td>
</tr>
<tr>
<td>Cyclist Behaviour/Personality</td>
<td>Average speed arterial</td>
<td>Average speed for each user on arterials</td>
</tr>
<tr>
<td></td>
<td>Average speed non-arterial</td>
<td>Average speed for each user on non-arterials</td>
</tr>
<tr>
<td></td>
<td>Average speed uphill</td>
<td>Average speed for each user on uphill segments</td>
</tr>
<tr>
<td></td>
<td>Average speed downhill</td>
<td>Average speed for each user on downhill segments</td>
</tr>
<tr>
<td></td>
<td>Average speed level</td>
<td>Average speed for each user on level segments</td>
</tr>
</tbody>
</table>
Table 6-3 shows the best model results for speed without considering any cyclist’s preferences. The R-squared value, representing the ability of the model to predict the speed, reaches a maximum of 0.12. This model identifies that cyclist speed is greater during work or school related trips, during morning peak and riding along arterials or collectors (versus local streets). Also, speeds are, on average, slower during the night. The longer the road segment, the more time a cyclist has to gain speed and therefore they go faster. Also, cyclists ride faster if the intersections at both ends of the segment are not signalized. This model also identifies that cyclists ride faster when the temperature is between 10 and 20 degrees Celsius compared to cold, below 10 degrees Celsius, and hot, over 20 degrees. Male cyclists and cyclists who ride in the winter, tend to go faster than female cyclists and those who don’t ride during Montreal’s cold and snowy winters. Cyclists 25 and younger tend to ride faster whereas cyclists older than 44 tend to ride slower.

Table 6-3 Comparison of Results using Unfiltered and Filtered Speeds

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median 7</th>
<th>Median 5</th>
<th>Median 3</th>
<th>Unfiltered</th>
<th>Median 7</th>
<th>Median 5</th>
<th>Median 3</th>
<th>Unfiltered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uphill</td>
<td>-4.680</td>
<td>0.084</td>
<td>0.000</td>
<td>-4.701</td>
<td>0.084</td>
<td>0.000</td>
<td>-4.730</td>
<td>0.085</td>
</tr>
<tr>
<td>Downhill</td>
<td>2.936</td>
<td>0.077</td>
<td>0.000</td>
<td>2.933</td>
<td>0.077</td>
<td>0.000</td>
<td>2.958</td>
<td>0.078</td>
</tr>
<tr>
<td>Arterial or collector</td>
<td>2.050</td>
<td>0.054</td>
<td>0.000</td>
<td>2.016</td>
<td>0.055</td>
<td>0.000</td>
<td>1.963</td>
<td>0.055</td>
</tr>
<tr>
<td>Work or school</td>
<td>0.701</td>
<td>0.053</td>
<td>0.000</td>
<td>0.691</td>
<td>0.053</td>
<td>0.000</td>
<td>0.686</td>
<td>0.054</td>
</tr>
<tr>
<td>Age under 25</td>
<td>1.254</td>
<td>0.111</td>
<td>0.000</td>
<td>1.236</td>
<td>0.112</td>
<td>0.000</td>
<td>1.173</td>
<td>0.113</td>
</tr>
<tr>
<td>Age above 44</td>
<td>-1.273</td>
<td>0.062</td>
<td>0.000</td>
<td>-1.274</td>
<td>0.062</td>
<td>0.000</td>
<td>-1.313</td>
<td>0.063</td>
</tr>
<tr>
<td>Length</td>
<td>0.016</td>
<td>0.000</td>
<td>0.000</td>
<td>0.016</td>
<td>0.000</td>
<td>0.000</td>
<td>0.016</td>
<td>0.000</td>
</tr>
<tr>
<td>a.m. peak</td>
<td>0.939</td>
<td>0.056</td>
<td>0.000</td>
<td>0.935</td>
<td>0.057</td>
<td>0.000</td>
<td>0.944</td>
<td>0.057</td>
</tr>
<tr>
<td>None signalized</td>
<td>1.085</td>
<td>0.054</td>
<td>0.000</td>
<td>1.060</td>
<td>0.054</td>
<td>0.000</td>
<td>1.015</td>
<td>0.055</td>
</tr>
<tr>
<td>Night</td>
<td>-0.300</td>
<td>0.060</td>
<td>0.000</td>
<td>-0.304</td>
<td>0.061</td>
<td>0.000</td>
<td>-0.296</td>
<td>0.062</td>
</tr>
<tr>
<td>Male</td>
<td>1.182</td>
<td>0.053</td>
<td>0.000</td>
<td>1.156</td>
<td>0.053</td>
<td>0.000</td>
<td>1.146</td>
<td>0.054</td>
</tr>
<tr>
<td>Winter</td>
<td>1.357</td>
<td>0.046</td>
<td>0.000</td>
<td>1.375</td>
<td>0.046</td>
<td>0.000</td>
<td>1.386</td>
<td>0.047</td>
</tr>
<tr>
<td>Bicycle infrastructure</td>
<td>0.408</td>
<td>0.046</td>
<td>0.000</td>
<td>0.413</td>
<td>0.047</td>
<td>0.000</td>
<td>0.422</td>
<td>0.047</td>
</tr>
<tr>
<td>Temperature 10° to 20°</td>
<td>0.170</td>
<td>0.047</td>
<td>0.000</td>
<td>0.153</td>
<td>0.047</td>
<td>0.001</td>
<td>0.121</td>
<td>0.048</td>
</tr>
<tr>
<td>Constant</td>
<td>13.199</td>
<td>0.111</td>
<td>0.000</td>
<td>13.330</td>
<td>0.111</td>
<td>0.000</td>
<td>13.537</td>
<td>0.113</td>
</tr>
<tr>
<td>Observations</td>
<td>88945</td>
<td>88945</td>
<td>88945</td>
<td>88945</td>
<td>88945</td>
<td>88945</td>
<td>88945</td>
<td>88945</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.120</td>
<td>0.118</td>
<td>0.116</td>
<td>0.106</td>
<td>0.116</td>
<td>0.106</td>
<td>0.106</td>
<td>0.106</td>
</tr>
</tbody>
</table>

Table 6-3 shows the model results using the unfiltered values, median of 3 values, median of 5 values and median of 7 values for speed. Table 6-3 shows that the parameter coefficients and
levels of significance are very similar but the model using the median of 7 values for speed performs best in terms of goodness-of-fit and modeling this dependent variable is preferred.

Overall, the goodness-of-fit, even of the best model, according to the R-squared, does not perform very well. In an attempt to improve the results, information of each cyclist’s preferences were extracted from the data. Speeds are not that easy to predict since cyclists who are the same age, gender, have the same trip purpose and use the same route can all travel at different speeds. In other words, two male cyclists of the same age and both riding to work along the same route, one can go slow and one go fast, and there are no variables that can capture personal preferences and differentiate between these two cyclists.

The following modeling results include variables which are specific to each cyclist. In other words, speed can be predicted for each individual cyclist based on his or her own behaviour and preferences. Especially for cyclists who used the smartphone application many times to track their trips, we can obtain average speeds along arterials and non-arterials, during peak hours and off-peak hours, when riding on segments uphill, on segments downhill or on level segments, during work or school trips and recreational or shopping trips, on hot days or on cold days or humid days and so on. When cyclists’ behaviour and personality are considered by adding one variable representing the average speed of each user along uphill, downhill and level segments, the R-squared value almost doubles to a value of 0.22. The model results are shown in Table 6-4.

<table>
<thead>
<tr>
<th>Table 6-4 Model Results with One Personality Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Average user speed on uphill, downhill and level</td>
</tr>
<tr>
<td>segments</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>R-squared</td>
</tr>
</tbody>
</table>

The personality model can be improved by adding some of the variables from Table 6-3. The results of the final model with all variables significant to the 95% level and without any issues of correlation between the independent variables, is shown in Table 6-5. To account for cyclist’s personality, the average speed on uphill, downhill and level segments as well as the average speed on arterials are included in the model. Therefore the dummy variables for slope and arterial are not included in this model. Once these personality variables are added, age and gender are no longer significant. Some of the application’s users opted out from providing their age and gender so once
these variables are removed from the model, we are able to model the speed on all segments since no information is missing. Adding the personality variables improved the fit of the speed model and all the parameter coefficients from the results in Table 6-3, which are also included in the model in Table 6-5, have reduced and the personality variables have a greater effect on speed.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work/school</td>
<td>0.253</td>
<td>0.045</td>
<td>0.000</td>
</tr>
<tr>
<td>Length</td>
<td>0.011</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>AM peak</td>
<td>0.263</td>
<td>0.050</td>
<td>0.000</td>
</tr>
<tr>
<td>Night</td>
<td>-0.214</td>
<td>0.052</td>
<td>0.000</td>
</tr>
<tr>
<td>Bicycle infrastructure</td>
<td>0.431</td>
<td>0.041</td>
<td>0.000</td>
</tr>
<tr>
<td>Both non-signalized</td>
<td>0.978</td>
<td>0.045</td>
<td>0.000</td>
</tr>
<tr>
<td>Temperature between 10° and 20°</td>
<td>0.169</td>
<td>0.041</td>
<td>0.000</td>
</tr>
<tr>
<td>User speed uphill, downhill and level</td>
<td>0.971</td>
<td>0.006</td>
<td>0.000</td>
</tr>
<tr>
<td>User speed on arterial</td>
<td>0.035</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.717</td>
<td>0.128</td>
<td>0.000</td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td>99344</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td></td>
<td>0.246</td>
<td></td>
</tr>
</tbody>
</table>

Flow variables do not have a significant effect on travel times. While one would expect speeds to be slower along roads with a high number of cyclists, this effect was not found to be significant in the speed models, likely because cyclist congestion is not yet an issue in Montreal.

Despite the effort to account for each cyclist’s personality, we are still unable to develop a strong model to predict cyclist speed along road segments. There are likely many personal attributes that we cannot account for.

6.6.4. Applications

One application of generating segment speeds and intersection delays is for route navigation. Routing algorithms can be updated with segment and intersection-specific travel speeds and times rather than assuming a constant speed for cyclists. Also, by accounting for cyclist’s preferences, the fastest route can be determined for each cyclist individually. The authors’ previous work carried out with this GPS data, mapped cyclist’s risk of injury throughout the entire network. Combining all this work, a tool can be developed for routing capable of identifying the most popular route (most used by cyclists in the data), the safest route (based on injury risk or hard braking), the fastest route (based on speeds and delays) and the shortest route (based on distance travelled). To show this tool, we have selected an origin and destination in the boroughs of Plateau Mont-Royal and
Ville-Marie respectively. The origin was selected as the Laurier Metro and the destination as Central Station. Using the road network combined with all the generated injury and speed data, a cyclist traveling from Laurier Metro to Central Station has been provided 3 routes as shown in Figure 6-6. This figure shows how different the routes are based on shortest route, fastest route and safest route.

Road geometry can also be accounted for as restrictions or preferences. For example, one restriction could be one-way streets, therefore not allowing the proposed routes to go the opposite way on one-way streets. For preferences, different levels of preference can be applied and in this work we looked at preference for cycle tracks. Three levels of preference, low, medium and high are the only options provided by ESRI's ArcCatalog when building the network. Figure 6-7 shows the resulting routes for the 3 cycle track preferences as well as the shortest route without any preference for cycle tracks.

This data can also be used to identify corridors with high cyclist volumes and very large delays which can push towards the implementation of green waves for cyclists. This means that cyclists traveling at the design speed along the road segments along the corridor should always arrive at the intersection during the green signal and therefore reducing their delays.
Figure 6-6 Shortest Route, Fastest Route and Safest Route

Figure 6-7 Shortest Routes Based on Different Levels of Preference for Cycle Tracks
6.7. CONCLUSION AND FUTURE WORK

Cyclist numbers are increasing and in response, cities are changing their approaches to infrastructure and signal phasing, but in order to propose appropriate design changes and signal phasing modifications to account for cyclists, reliable travel times, speeds, delays and level-of-service measures are necessary for cyclists. This work presented a methodology to clean GPS data and compute speeds along segments and delays through intersections using second-by-second cyclist GPS data from a Smartphone application in Montreal. The main benefit of using GPS data over traditional data collection methods is that, GPS data provides the potential of capturing the trip routes disaggregated to second-by-second locations for a large sample of cyclists over the entire road and bicycle facility network and not only at specific points in the network. GPS data also provides the actual route taken by each cyclist from their origin to their destination. In this work, we explored the use of recent GPS cyclist trip data, from a Smartphone application, for identifying different level-of-service measures such as travel time, speed and delay at the level of the entire road and intersection network for the island of Montreal. Until now, very little has been known about cyclist speeds and delays at the disaggregate level of each road segment and intersection which is vital information required by cities to provide better level-of-service for cyclists. This data can also be used to provide cyclists with accurate travel times for routing purposes.

The speeds along all segments, in both directions, in downtown Montreal were computed and mapped. On average cyclist speed is greater along arterials than on local streets and greater along bicycle infrastructure than along segments without any. The developed methodology was applied to compute cyclist delays through intersections based on the difference in time to travel through the intersection traveling at segment speed versus traveling at intersection speed. These results were also mapped showing the delay on each approach as well as the average per intersection. These speeds and delays serve as vital information for navigation and routing purposes since they can identify speeds and delays during different times of the day and how they differ across roads and bicycle facilities to provide more accurate travel times and shortest path routes when determining directions from origin to destination. Knowledge of speeds and delays can also lead to further analysis to study and reveal areas where there may be bottlenecking or geometric design issues leading to actual speeds and delays differing from their ideal values.

Modeling cyclist speed revealed that the variable representing the cyclists’ average speed on uphill, downhill and level segments, alone improves the fit of the model drastically. Adding the
cyclists’ average speed on arterials as well as geometric design and built environment attributes further improved the model reaching an R-squared value of 0.246. The segment speed model results showed that personal preferences and behaviour are the most important factors affecting cyclist speed, in addition to some geometric design and built environment characteristics. The model results also identify that segments which have cyclists biking for work or school related purposes, segments used during morning peak, segments with bicycle infrastructure and segments which do not have signalized intersections at either end, tend to have cyclists riding at greater speeds. Also, cyclists travel faster when the temperature is between 10° and 20° Celsius and travel slower late at night or early morning.

As part of future work, different model formulations for the speed along segments will be tested and additional variables will be included. Also, the above work will be updated using the newest GPS data since the application continues to collect cyclist trip data and the methodology will be applied to the entire island of Montreal provided there is enough trip data. Also with more trip data, the delays can be computed during peak and off peak periods to see if speeds and delays change and by how much.

The most important goal of our future work is to combine all this data about delay, speed, travel time and LOS, to study their effect on cyclist safety along segments and at intersections and further to see how these affect the safety, delay and speed of other modes (vehicles and pedestrians) sharing the same segments and intersections. As part of future work, delays for all cyclist movements (through as well as right and left turns) will be computed and the delays will be disaggregated further by trip purpose: work/school related or not. Also, delays will be obtained using different methods based on speed and acceleration/deceleration profiles. For example, we can identify when each cyclist begins to slow down as they approach an intersection followed by the time that they are stopped and when they begin to accelerate and reach cruising speeds once more. Also, different model formulations for speed will be tested and further analysis will be carried out to determine what factors affect cyclist speeds on segments.

6.8. ACKNOWLEDGMENTS

We acknowledge the financial support provided by the FRQNT and NSERC. We would like to thank the Montreal Department of Public Health and Urgences-santé for collecting and validating
the injury data as well as Brisk Synergies and the city of Montreal for the GPS data. All remaining errors and the views expressed in this research are, however, solely ours.

6.9. REFERENCES


Chapter 7

Conclusion and Future Work
Chapter 7: Conclusion and Future Work

7.1. GENERAL CONCLUSION

Bicycle activity in North America is experiencing a rising trend but as bicycle flows increase so does the likelihood that cyclist injuries will occur. Despite the extensive literature on cyclist injury occurrence, few studies have focused on city-level safety modeling tools for cyclists. Also, few studies have looked at the various contributing factors associated with cyclist injury occurrence across different facility types. Methods for mapping bicycle volumes and safety in the entire network are missing, using both traditional and surrogate safety approaches. To deal with the endogeneity problem, statistical approaches to model bicycle flows and safety outcomes were developed to address the gaps in the literature. For urban areas, multimodal safety studies in the literature are also very rare. This work proposed new methodologies to simultaneously compute cyclist injury occurrence and bicycle activity models at intersections, integrating different sources of information such as short and long-term counts using automated counters as well as GPS data for cyclist trips. Some key contributions include the integration of methodologies considering a multimodal approach and multiple facility types. In addition, the proposed methods start from estimating activity and injury risk at specific sites and end at studying these at the entire network level. The proposed and developed methodologies are applied to the entire road network on the island of Montreal and empirical evidences are documented as part of this thesis.

More specifically, Chapter 2 expanded and improved upon the existing literature by proposing a simultaneous modeling approach to integrate bicycle activity and injury occurrence at signalized intersections. A Bayesian framework was proposed for modeling these two outcomes simultaneously. This work identified the key contributing factors associated with cyclist safety and bicycle activity at intersections. Using Bayesian analysis, expected injuries and injury rates were computed and used to rank cycling corridors and identify the most dangerous corridors on the island of Montreal. Compared to the current literature, this work studied a large sample of intersections for which both cyclist and motor-vehicle counts were collected at a disaggregate level as well as many geometric design and built environment characteristics. The modeling approach used in this work also accounted for the potential presence of endogeneity and unobserved heterogeneities. This chapter identified to what degree cyclist injury occurrence increases with
increasing traffic volumes and also identified that the presence of bus stops increases cyclist injury occurrence. This chapter also identified the effects that other geometric design and built environment characteristics have on cyclist injury and bicycle activity outcomes.

Chapter 3 proposed an extension of the methodology developed in Chapter 2 to investigate safety at intersections from a multimodal approach, considering all road users who share the same facilities, namely cyclists, pedestrians and motor-vehicles. This chapter, similar to Chapter 2, focused on signalized intersections but in addition, considered safety at non-signalized intersections. This chapter studied and revealed the effects of geometric design and built environment characteristics on the injury occurrence of cyclists, pedestrians and motor-vehicle occupants at signalized and non-signalized intersections. An inventory of data, including multimodal flow data, geometric design and built environment characteristics, were all collected for large samples of signalized and non-signalized intersections and used in the analysis. The multimodal safety results emphasize the importance of motor-vehicle flow on both cyclist and pedestrian injury occurrence. The outcomes also emphasize the effect of motor-vehicle flow on cyclist and pedestrian risk at intersections. At signalized intersections, the risk that vehicles impose on cyclists and pedestrians is on average, 14 and 12 times greater than for motor-vehicle occupants, respectively.

Chapter 4 addressed the importance of exposure measures and developed a methodology to obtain average annual daily bicycle flows throughout the entire network of road segments and intersections by combining bicycle counts and cyclist GPS data. More specifically, the developed methodology combines short and long-term data from manual and automatic counts, respectively, to develop an extrapolation function to adjust the GPS counts to average annual daily bicycle values using Montreal as a case study. The previous chapters (Chapters 2 and 3) only studied safety at a sample of intersections since count data was only available for these sites. With data from the Smartphone application for cyclist trips, AADB volumes are not only available for intersections but also for segments and for the entire population of sites for both, so safety applications were carried out at the level of the entire road network. Flows and injury risk were mapped and comparative analysis was carried out. Among other results, signalized intersections, which are often located at the intersection of major arterials, experience 4 times more injuries and 2.5 times greater risk than non-signalized intersections. A similar observation can be made for arterials which not only have a higher concentration of injuries but also injury rates (risk). On average, streets with
cycle tracks have a greater concentration of injuries due to greater bicycle volumes, however, and in accordance with recent work, the individual risk per cyclist is lower, justifying the benefits of cycle tracks.

Chapter 5 proposed a surrogate safety method based on the Smartphone GPS data. This was proposed as an alternative or complementary network screening method which can be used when accident data is unavailable or is incomplete, but also to identify dangerous locations for cyclists without the need to wait for accidents to occur. As a first step, a surrogate safety measure for cyclists based on deceleration rate was computed. Then, for validation purposes, the use of deceleration rate as a surrogate measure was compared to observed injuries. Using the Spearman rank correlation coefficient, the ranking of sites based on the expected number of injuries using the Empirical Bayes estimator (EB) and based on dangerous decelerations were compared. A relatively high correlation was found to exist between the proposed surrogate safety measure and expected injuries. The ranks of injuries and dangerous decelerations have a correlation of 0.55 at signalized intersections, 0.37 at non-signalized intersections and 0.48 at segments. Isolating intersections and segments in the central business district of Montreal, the coefficient stays the same for signalized intersections and increases to 0.55 for non-signalized intersections and to 0.58 for segments. These results show that deceleration rate, obtained from GPS data, can enable a shift from a reactive to a proactive approach and can hopefully prevent injuries due to identifying and treating dangerous locations prior to witnessing injuries and even fatalities.

Chapter 6 extended the applications of the Smartphone GPS data from safety applications to mobility applications. In this chapter, a methodology was developed to compute and map cyclist speeds along road segments and delays through signalized intersections in the network. The results found that cyclist speeds are greater along arterials and when riding on cycle tracks. This chapter went further and combined network risk measures with distances, delays and travel times, to be used for routing purposes and to be integrating into routing tools. With this information, a routing algorithm is capable of providing routes for cyclists based not only on real travel times, accounting for intersection delays, or shortest distances but also based on safety indicators, such as injury risk and surrogate safety measures. For instance, if a cyclist wants to travel along the safest route from their origin to destination, an algorithm could be created to determine their ideal route by minimizing their risk based on injury data or by minimizing risk based on surrogate safety measures. Depending on each cyclist's preferences, routes can be provided based on minimizing
more than one measure by weighing their preferences. This last chapter showed the importance of combining traditional safety measures and surrogate safety measures with mobility measures for routing purposes.

Overall, this thesis contributes to the existing literature in various ways, by: i) proposing a simultaneous Bayesian modeling approach to investigate cyclist injury occurrence and bicycle activity at intersections, ii) presenting a multimodal framework for analyzing safety across modes at signalized and non-signalized intersections, iii) developing a data-fusion based approach to map bicycle volumes in the entire network combining short- and long-term count data with GPS data from a Smartphone application, iv) proposing a surrogate safety measure for network screening based on decelerations from GPS data, and v) extracting other network performance measures such as travel speeds and delays in order to propose a routing (navigation) tool that combines multiple safety and mobility criteria.

7.2. FUTURE WORK

As part of future work, the methodologies proposed in this thesis could be applied and validated using data from other cities. This will help investigate the issues of transferability. Some of the processes, such as data integration, development of safety performance functions and the mapping of cyclist risk could be automated in order to facilitate the process of applying the methodology to other cities.

With respect to surrogate safety measures, the proposed measure based on decelerations needs more validation using more data from the same city and other cities. Also, a separate tool can be developed to integrate a routing algorithm with all the proposed safety and mobility measures. This algorithm will be designed to be capable of applying each cyclist’s preferences for what attribute or attributes they wish to minimize for each trip, whether they wish to minimize travel time based on speeds along segments and delays through intersections, distance travelled or risk of injury through intersections and along segments as well as any combination of measures.

In order to validate the predicted average annual daily bicycle flows at non-signalized intersections and along road segments, a larger sample of manual, automatic or any other source of count data is required for both sites. As the next step, it is important to define the optimal amount of data necessary, such as the optimal number of sites with manual short-term and automatic long-
term counts as well as the minimum number of GPS trips required to carry out and repeat this analysis for other cities.

Also, we would like to combine the safety results with the mobility results to see how they relate to one another and answer research questions related to the effects that delays have on cyclist safety at intersections to identify whether or not delays are good or bad from a safety perspective. Ideally, we would like to obtain similar mobility data for vehicles and pedestrians and look at the relationship between the safety and mobility of all modes simultaneously to see how delays for one mode affect safety for other modes or study the effect that providing safer roads for cyclists may have on vehicle delays. As a continuation of this work, we would also like to test different buffer dimensions and dangerous deceleration thresholds to test the sensitivity of the surrogate safety measure results. For cyclist delays at intersections, turning movements in addition to through movements should also be integrated.

7.3. PUBLICATIONS

Refereed journal and conference papers in chronological order from newest to oldest.


   - Accepted for presentation at the Transportation Research Board 95th Annual Meeting, Washington D.C., USA 2016.

   - Presented at the 25th Canadian Association of Road Safety Professionals Conference (previously called the Canadian Multidisciplinary Road Safety Conference), Ottawa, Ontario, 2015.
   - Presented at the Transportation Research Board 94th Annual Meeting, Washington D.C., USA 2015
   - Presented at the 25th Canadian Association of Road Safety Professionals Conference (previously called the Canadian Multidisciplinary Road Safety Conference), Ottawa, Ontario, 2015.


   - Presented at Transportation Research Board 93rd Annual Meeting, Washington D.C., USA 2014.

   - Presented at the Transportation Research Board 92nd Annual Meeting, Washington D.C., USA 2013.

   - Presented at Transportation Research Board 91st Annual Meeting, Washington D.C., USA 2012.

    - Presented at the 23rd Canadian Multidisciplinary Road Safety Conference, Montreal, Quebec, 2013.
    - Presented at the Canadian Society for Civil Engineering General Conference, Montreal, Quebec, 2013.
   ▪ Presented at the 22nd Canadian Multidisciplinary Road Safety Conference, Banff, Alberta.
   ▪ Presented at the Association Québécoise du Transport et des Routes in Quebec City, Quebec, 2012.


   ▪ Presented at McGill Civil Engineering Graduate Students Society Conference, 2011.


Appendix A - Non-Signalized Intersections

The paper titled, Bayesian Modeling Approach for Cyclist Injury Risk: Comparison of Signalized and Non-Signalized Intersections, was presented at the Canadian Multidisciplinary Road Safety Conference in 2013 and was awarded the second place student paper by the Insurance Bureau of Canada (IBC). This paper was also presented at the Canadian Society for Civil Engineering General Conference in 2013. This paper studies cyclist injury risk as in Chapter 2 at signalized intersections and at non-signalized intersections as shown in this Appendix. The simultaneous injury and activity results for non-signalized intersections have not been reported in Chapter 2 but the site selection and description of the data at non-signalized intersections have been discussed in Chapter 3.

Here are the results for the simultaneous Bayesian analysis of cyclist injury and bicycle activity at non-signalized intersections.

A.1. RESULTS

A.1.1. Factors Associated with Cyclist Injury Risk

Consistent with the literature, cyclist injury results reveal the importance of both bicycle and motor-vehicle flows on cyclist injury occurrence (Table A-1a). A 1.0% increase is bicycle flows is expected to cause a 0.75% increase in cyclist injury occurrence. At non-signalized intersections disaggregate motor-vehicle movements were not found to have significant effects on injury occurrence. A 1.0% increase in total motor-vehicle flows however is expected to increase injury occurrence by 0.39%. Of the intersection attributes tested, only the total number of lanes entering the intersection was found to have a significant effect on injury occurrence. Adding a lane of traffic would cause cyclist injuries to increase by 15.5%.

A.1.2. Factors Associated with Bicycle Activity

The results for the bicycle activity model is shown in Table A-1b. This study assumed that bicycle activity through non-signalized intersections can be predicted using the same variables as activity through signalized intersections. All the variables found to be significant for signalized intersections were also found to be significant for non-signalized intersections. This hypothesis
requires more work to test whether there are different variables that should be considered to better predict bicycle activity through non-signalized intersections.

These results also reveal that the covariance term between the injury and activity models was found to be insignificant. This may be due to a smaller sample size as well as much fewer injuries observed at these locations. Additional work is required.

**Table A-1 Non-Signalized Intersection Results**

<table>
<thead>
<tr>
<th>a) Cyclist Injury Model</th>
<th>Coefficient</th>
<th>Standard Deviation</th>
<th>95% Credible Interval</th>
<th>Elasticity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln* bicycle flows</td>
<td>0.754</td>
<td>0.102</td>
<td>0.525</td>
<td>0.919</td>
</tr>
<tr>
<td>Ln* total motor-vehicle flows</td>
<td>0.399</td>
<td>0.115</td>
<td>0.204</td>
<td>0.622</td>
</tr>
<tr>
<td>Total number of lanes</td>
<td>0.168</td>
<td>0.102</td>
<td>-0.032</td>
<td>0.349</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.526</td>
<td>1.756</td>
<td>-6.26</td>
<td>-0.73</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b) Bicycle Activity Model</th>
<th>Coefficient</th>
<th>Standard Deviation</th>
<th>95% Credible Interval</th>
<th>Elasticity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>400m Employment ('000)^2</td>
<td>0.072</td>
<td>0.002</td>
<td>0.067</td>
<td>0.076</td>
</tr>
<tr>
<td>400m Presence of schools^2</td>
<td>0.178</td>
<td>0.043</td>
<td>0.103</td>
<td>0.242</td>
</tr>
<tr>
<td>800m Metro (subway) stations^3</td>
<td>0.214</td>
<td>0.005</td>
<td>0.204</td>
<td>0.222</td>
</tr>
<tr>
<td>800m Land use mix^3</td>
<td>0.379</td>
<td>0.042</td>
<td>0.313</td>
<td>0.502</td>
</tr>
<tr>
<td>800m Length of bicycle facilities^3</td>
<td>0.181</td>
<td>0.005</td>
<td>0.173</td>
<td>0.190</td>
</tr>
<tr>
<td>50m Area of commercial land use ('000)^1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Presence of three approaches</td>
<td>-0.229</td>
<td>0.024</td>
<td>-0.260</td>
<td>-0.173</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.293</td>
<td>0.020</td>
<td>-0.332</td>
<td>-0.256</td>
</tr>
<tr>
<td>DIC</td>
<td></td>
<td></td>
<td>3456</td>
<td></td>
</tr>
<tr>
<td>Covariance</td>
<td></td>
<td></td>
<td>0.385</td>
<td></td>
</tr>
</tbody>
</table>

*Ln = natural logarithm

1-3 - These variables were measured within a 1) 50 m, 2) 400 m or 3) 800 m radius around the intersections.
Appendix B - Additional Multimodal Results

The versions of this paper presented at the 24th Canadian Multidisciplinary Road Safety Conference, 2014 and at the Transportation Research Board 93rd Annual Meeting, 2014, showed and discussed different results from the ones published in Accident Analysis and Prevention but remain interesting to include in this thesis. This appendix presents those results.

B.1. EXPECTED NUMBER OF INJURIES AND RISK

The graphs in Figure B-1 show the expected number of injuries and risk at signalized and non-signalized intersections for the three modes of interest. These graphs are intended to show the overall distribution of injuries and risk and are plotted after sorting injuries or risk based on the best to the worst intersections for cyclists, pedestrians and motor-vehicles separately.

At signalized intersections, the expected number of pedestrian injuries is greater than that of cyclists, as can be seen in Figure B-1 (top left). This figure also shows that injuries occurring to drivers and passengers of motor-vehicles are far greater, in absolute numbers, than either form of active transportation. Referring to Figure B-1 (top right), at the lower levels of risk for signalized intersections in this sample (below 2 injuries per million cyclists, or pedestrians), the risk is higher for cyclists. In other words, for approximately 420 of the 647 intersections (65%), the risk is higher for cyclists than pedestrians. This graph also shows that at high levels of risk, the risk becomes even greater for pedestrians than cyclists. When accounting for the exposure of motor-vehicles, even though there are many injuries, they still have by far the lowest risk. Also, looking at the difference in the shape of the expected number of injuries (green curve in Figure B-1 top left) and risk (green curve in Figure B-1 top right) curves for motor-vehicles, we see that the injury curve increases and covers a range from 0 to 25 injuries. The risk curve on the other hand is relatively horizontal. Together these curves reveal that locations where injury frequency is high, are also locations where traffic flow is high and therefore the risk for motor-vehicles at signalized intersections are quite uniform and relatively low in comparison to other modes.

At the non-signalized intersections in this sample, the expected number of injuries for cyclists, pedestrians and motor-vehicles are much lower than at signalized intersections and all three model curves follow very similar patterns. In this case, the expected number of cyclist
injuries is greater than that of pedestrians (Figure B-1 bottom left). Once again, as observed for signalized intersections, there are more motor-vehicle than cyclist and pedestrian injuries. The difference between motor-vehicle and cyclist and pedestrian injuries however is much smaller at non-signalized than at signalized intersections. At non-signalized intersections, cyclists are exposed to much greater risks of injury than pedestrians and motor-vehicles (Figure B-1 bottom right). While the injury curves are very similar, the risk curves at non-signalized intersections demonstrate a very different pattern. Over the entire sample of non-signalized intersections, the risk for pedestrians, from minimum to maximum, only reaches a value slightly greater than 0.5 (similar for motor-vehicles) whereas cyclist risk almost reaches a value of three.
Figure B-1 Number of Injuries and Risk at Signalized (a and b) and Non-Signalized (c and d) Intersections
B.2 CORRELATIONS BETWEEN FLOWS, INJURIES AND RISK

An additional step was carried out to test whether there exists correlation between cyclist, pedestrian and motor-vehicle flows, injuries and risk. Knowing if and where there is correlation between these variables, can provide guidance for where to implement safety improvements and useful information about each site. For example, if an intersection identified as being dangerous for cyclists is also identified as being dangerous for pedestrians, safety improvements can be carried out for both modes at once. On the one hand this situation can reveal some design flaws for those particular intersections and on the other hand this can ease the task of selecting sites for interventions as well as potentially provide a more economically viable solution for the same level of safety improvement (compared to the cost of improving two separate sites: one dangerous for cyclists and one dangerous for pedestrians).

Table B-1 reiterates, as did the injury models, that cyclist injuries are correlated with bicycle flows, pedestrian injuries with pedestrian flows and motor-vehicle with motor-vehicle flows. Table B-1 also reveals that for both signalized and non-signalized intersections, cyclist and pedestrian flows are correlated implying that intersections with high bicycle flows also have high pedestrian flows. For signalized intersections there is a negative relationship between both bicycle and pedestrian flows with respect to motor-vehicle flows. This means that non-motorized users are traveling where levels of motorized traffic is lower. At non-signalized intersections this is not the case. However this result may be due to the lower levels of traffic passing through these facilities. Also at signalized intersections but not at non-signalized intersections, cyclist risk is correlated with both pedestrian and motor-vehicle risk. Also, vehicle injury is highly correlated with not only vehicle risk but also with cyclist and pedestrian risk at both facilities.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Signalized Intersections</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bicycle Flow</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Flow</td>
<td>0.28</td>
<td>1</td>
<td></td>
<td>0.19</td>
<td>0.11</td>
<td>0.30</td>
<td>0.37</td>
<td>0.21</td>
<td>0.12</td>
</tr>
<tr>
<td>Vehicle Flow</td>
<td>-0.13</td>
<td>-0.15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bicycle Injury</td>
<td>0.31</td>
<td>0.19</td>
<td>0.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Injury</td>
<td>0.16</td>
<td>0.35</td>
<td>0.11</td>
<td>0.30</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle Injury</td>
<td>-0.03</td>
<td>-0.05</td>
<td>0.25</td>
<td>0.15</td>
<td>0.19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclist Risk</td>
<td>-0.19</td>
<td>-0.07</td>
<td>0.08</td>
<td>0.04</td>
<td>0.01</td>
<td>0.37</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Risk</td>
<td>-0.13</td>
<td>-0.14</td>
<td>0.18</td>
<td>-0.05</td>
<td>0.02</td>
<td>0.21</td>
<td>0.35</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Vehicle Risk</td>
<td>0.03</td>
<td>0.01</td>
<td>-0.04</td>
<td>0.13</td>
<td>0.19</td>
<td>0.85</td>
<td>0.22</td>
<td>0.12</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-Signalized Intersections</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bicycle Flow</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Flow</td>
<td>0.36</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle Flow</td>
<td>0.25</td>
<td>0.16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bicycle Injury</td>
<td>0.39</td>
<td>0.21</td>
<td>0.12</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Injury</td>
<td>0.16</td>
<td>0.25</td>
<td>0.15</td>
<td>0.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle Injury</td>
<td>0.13</td>
<td>0.11</td>
<td>0.34</td>
<td>0.15</td>
<td>0.25</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclist Risk</td>
<td>-0.12</td>
<td>-0.05</td>
<td>0.12</td>
<td>0.35</td>
<td>-0.02</td>
<td>0.13</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedestrian Risk</td>
<td>-0.04</td>
<td>-0.21</td>
<td>0.45</td>
<td>-0.06</td>
<td>0.27</td>
<td>0.21</td>
<td>0.06</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Vehicle Risk</td>
<td>-0.03</td>
<td>0.02</td>
<td>-0.08</td>
<td>0.04</td>
<td>0.21</td>
<td>0.81</td>
<td>0.03</td>
<td>0.03</td>
<td>1</td>
</tr>
</tbody>
</table>
Appendix C - Map-Matching, Additional Safety Application and GPS Data

C.1 MAP-MATCHING

As mentioned, after publishing this work, the authors implemented a map-matching algorithm, TrackMatching (Marchal, 2015), to snap the raw GPS data onto the road network using OpenStreets, to validate the simple buffer approach used to associate GPS points to intersections and segments in the network in Chapter 4. This tool however is proprietary and the specific method used to map-match is not known to the user but it is safe to assume that the algorithm considers the points before and after the current data point to consider the route when map-matched as a better way of identifying where the point should be snapped on the network. After applying this map-matching algorithm to the GPS data, a comparison between the map-matched data and the raw GPS data associated to the road and intersection networks based on the simple buffer approach was carried out. The results in Table C-1 show the results of the Spearman's rank correlation coefficient for each of the three facility types separately. According to this table, the simple buffer approach works well enough and is best at signalized intersections. Also, it is worth mentioning that even the map-matching algorithm is not always capable of snapping the points onto the network with high levels of precision. While the map-matching algorithm does provide an error value, unfortunately no explanation of its meaning is provided.

Table C-1 Correlation between Raw and Map-Matched GPS Counts for Different Facility Types

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Correlation between Raw and Map-Matched GPS Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signalized Intersection</td>
<td>0.8731</td>
</tr>
<tr>
<td>Non-Signalized Intersection</td>
<td>0.8374</td>
</tr>
<tr>
<td>Road Segment</td>
<td>0.8085</td>
</tr>
</tbody>
</table>

C.2 INJURY MAPS

In Chapter 4, the Smartphone GPS data was used to map bicycle flows and injury risk in the network. Using the empirical Bayes estimator the expected number of injuries can also be
computed and mapped throughout all intersections and road segments on the island. Figure C-1 shows maps of the expected number of injuries side by side with maps of expected risk of injury. Figure C-1 identifies that for both signalized and non-signalized intersections, cyclist injury occurrence is greater in the central neighborhoods of the island. However, when flows are accounted for in the measure of risk, we see an opposite pattern to injuries, and cyclist risk of injury is greater outside the central neighbourhoods. The maps for segments do not seem to follow the same trend. This is likely because overall there are so few cyclist injuries along segments but also because all injuries occurring along segments occurred at only 4% of all segments on the island. This means that 96% of segments (over 42,000) did not witness any cyclist injuries during the study period. Also, at intersections with high vehicle traffic, there are fewer gaps for cyclists to violate the traffic signal or stop sign, meaning that the chance of them getting in an accident is much lower. This can explain the lower risk at downtown intersections. At segments, for injuries to occur, there needs to be vehicles and segments downtown have far more vehicle traffic than outside of downtown which can also explain this discrepancy. Also even though cyclist risk at segments is high in the central neighbourhoods, overall the risk is by far lower than at intersections. Even though risk may be the highest in downtown, this value is still very low so likely not even that dangerous in comparison to intersections for which the risk values are much greater.
Figure C-1 EB Injuries and Risk at Signalized Intersections (Top), Non-Signalized Intersections (Middle) and Segments (Bottom)
C.3 SAFETY APPLICATION

Based on EB injuries and risk, the top 25 signalized intersections, non-signalized intersections and road segments, based on each criterion, can be identified. These sites are shown in Figure C-2. As we already saw in Figure C-1, the risk at intersections is greatest outside the central neighbourhoods of the city. More than half of these sites (about 60%) are located on the West end of the island. The West end of the island has 3 main North-South running boulevards for which, especially for des Sources, have the most dangerous intersections. On des Sources, 10 out of the 25 signalized intersection hotspots have been identified and two on Boulevard Saint Jean. The West end of the island has several cycle tracks running East to West but no North to South tracks to connect them. Also, some intersections either at one end or in the middle of segments with cycle tracks ranked high in terms of cyclist risk at intersection on the West and East ends but not in the central neighbourhoods. The remaining 40% of dangerous intersections are located in the East end, mostly in the Pointe-aux-Trembles-Rivières-des-Prairies borough.

For segments, eight of the 25 most dangerous segments were identified to be on Ontario and two on President Kennedy which is the continuation of Ontario to its West. Just one block South of Ontario, is de Maisonneuve, which has a cycle track running parallel to Ontario. Despite the cycle track, many cyclists chose to ride on Ontario. Based on risk, this should be avoided. Knowledge of where risk is highest for both intersections and segments, as was revealed in this study, makes it possible to propose alternate routes for cyclists and encourage the use of safer routes and also provides vital information to cities seeking to select sites for treatments.

In terms of injuries along segments, the majority of the hotspots are located within the central neighbourhoods. While injuries and risk are very low or even zero on cyclist tracks, the intersections along cycle tracks in Montreal rank very high in terms of cyclist injury occurrence. For segments, some which ranked high in terms of risk also seem to rank high in terms of injuries, so further analysis is warranted at these sites.
Figure C-2 Top 25 Most Dangerous Signalized Intersections, Non-Signalized Intersections and Segments Based on EB Injuries (Top) and EB Risk (Bottom)
The Smartphone GPS data was not thoroughly explained in Chapters 4 to 6 due to space restrictions of the journals and conferences these papers were submitted to. A more in-depth description of the GPS data, as well as more information about the long-term count data is provided in this appendix.

C.4 SMARTPHONE GPS TRIPS AND TRACES

In addition to latitudes, longitudes and timestamps, the Mon RésoVélo Smartphone application also collects trip and user attributes for each trip. The trip attributes include, origin and destination, trip start and end times and trip purpose. The user attributes include, age, gender, income and well as cycling habits, ridership frequency, experience and whether or not they bike during the winter. A summary of the users is provided in Table C-2. Table C-2 shows that the 25 to 34 age group is most represented, representing 42% of the application’s users. The next age group, 35 to 44 represents about 26% of the users. These values are obtained from the 721 of the 900 users who provided their age. Also, 73% of the users are male. Only 60% of the users provided their income range. According to the statistics in Table C-2, 27% of cyclists fall inside the highest income bracket whereas the remaining 73% of users are split fairly evenly throughout the other five categories. 45% of cyclists appear to bike daily and 41% appear to bike several times per week. Over 69% of the cyclists say that they have been cycling since childhood and almost 24% bike during the winter.

The study period is 137 days long from July 2nd, 2013 to November 15th, 2013 and in this short time, over 10,000 trips were recorded by about 900 cyclists, representing over 16 million GPS data points needing to be processed to meet our objectives. Figure C-3 shows the daily and cumulative number of cyclists registered in the application and Figure C-4 shows the daily and cumulative number of trips over the 137 days and Figure C-5 shows a sample of raw GPS points and traces.
## Table C-2 Summary Statistics of Trip and User Attributes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Categories</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group</strong></td>
<td>&lt; 18</td>
<td>5</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>18-24</td>
<td>69</td>
<td>9.57</td>
</tr>
<tr>
<td></td>
<td>25-34</td>
<td>306</td>
<td>42.44</td>
</tr>
<tr>
<td></td>
<td>35-44</td>
<td>186</td>
<td>25.8</td>
</tr>
<tr>
<td></td>
<td>45-54</td>
<td>96</td>
<td>13.31</td>
</tr>
<tr>
<td></td>
<td>55-64</td>
<td>51</td>
<td>7.07</td>
</tr>
<tr>
<td></td>
<td>65 and up</td>
<td>8</td>
<td>1.11</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td>721</td>
<td>100</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>Female</td>
<td>190</td>
<td>26.72</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>521</td>
<td>73.28</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td>711</td>
<td>100</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td>Less than $20,000</td>
<td>74</td>
<td>13.6</td>
</tr>
<tr>
<td></td>
<td>$20,000-$39,999</td>
<td>81</td>
<td>14.89</td>
</tr>
<tr>
<td></td>
<td>$40,000-$59,999</td>
<td>95</td>
<td>17.46</td>
</tr>
<tr>
<td></td>
<td>$60,000-$74,999</td>
<td>64</td>
<td>11.76</td>
</tr>
<tr>
<td></td>
<td>$75,000-$99,999</td>
<td>84</td>
<td>15.44</td>
</tr>
<tr>
<td></td>
<td>$100,000 and greater</td>
<td>146</td>
<td>26.84</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td>544</td>
<td>100</td>
</tr>
<tr>
<td><strong>Cycling frequency</strong></td>
<td>Less than once a month</td>
<td>4</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>Several times per month</td>
<td>47</td>
<td>12.98</td>
</tr>
<tr>
<td></td>
<td>Several times per week</td>
<td>148</td>
<td>40.88</td>
</tr>
<tr>
<td></td>
<td>Daily</td>
<td>163</td>
<td>45.03</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td>362</td>
<td>100</td>
</tr>
<tr>
<td><strong>Experience</strong></td>
<td>Since childhood</td>
<td>469</td>
<td>69.17</td>
</tr>
<tr>
<td></td>
<td>Several years</td>
<td>151</td>
<td>22.27</td>
</tr>
<tr>
<td></td>
<td>One year or less</td>
<td>37</td>
<td>5.46</td>
</tr>
<tr>
<td></td>
<td>Just started</td>
<td>21</td>
<td>3.1</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td>678</td>
<td>100</td>
</tr>
<tr>
<td><strong>Winter cycling</strong></td>
<td>No</td>
<td>518</td>
<td>76.63</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>158</td>
<td>23.37</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td>676</td>
<td>100</td>
</tr>
</tbody>
</table>
Figure C-3 Daily and Cumulative Number of Cyclists Registered

Figure C-4 Daily and Cumulative Number of Trips
Figure C-5 Raw GPS Points (left) and Traces (right)