Monitoring changes in patterns of cycling safety and ridership:
A spatial analysis

by

Darren George Boss
B.Sc., University of Victoria, 2011

SUPERVISORY COMMITTEE:

Dr. Trisalyn A. Nelson, Supervisor
(Department of Geography, University of Victoria)

Dr. Meghan Winters, Member
(Department of Geography, University of Victoria)
ABSTRACT

Cycling is an underutilized mode of transportation in cities across North America. Numerous factors contribute to low ridership levels, but a key deterrent to cycling is concern for personal safety. In an effort to increase cycling mode share, many cities are investing in cycling infrastructure, with several cities constructing connected bicycle networks. Monitoring the impact of new infrastructure is important for accountability to citizens and to encourage political will for future investments in cycling facilities. A lack of spatially continuous ridership data and methodological challenges have limited monitoring and evaluation of the impacts of infrastructure changes. The goal of our research was to demonstrate spatially explicit approaches for monitoring city-wide changes in patterns of safety and ridership following improvements to cycling infrastructure.

To meet our goal, our first analysis demonstrated a method for monitoring changes in the spatial-temporal distribution of cycling incidents across a city. We compared planar versus network constrained kernel density estimation for visualizing cycling incident intensity across the street network of Vancouver, Canada using cycling incidents reported to the Insurance Corporation of British Columbia. Next, we applied a change detection algorithm to detect statistically significant change between maps of kernel density estimates. The utility of the network kernel density change detection method is demonstrated through a case study in the city of Vancouver, Canada where we compare cycling incident densities following construction of two cycle tracks in the downtown core. The methods developed and demonstrated for this study provide city planners, transportation engineers and researchers a means of monitoring city-wide
changes in the patterns of cycling incidents following enhancements to cycling infrastructure.

Our second analysis demonstrated how network constrained spatial analysis methods can be applied to emerging sources of crowdsourced cycling data to monitor city-wide changes in patterns of ridership. We used network constrained global and local measures of spatial autocorrelation, applied to crowdsourced ridership data from Strava, to examine changes in ridership patterns across Ottawa-Gatineau, Canada, following installation and closures of cycling infrastructure. City planners, transportation engineers and researchers can use the methods outlined here to monitor city-wide changes in ridership patterns following investment in cycling infrastructure or other changes to the transportation network.

Through this thesis we help overcome the challenges associated with monitoring the impact of infrastructure changes on ridership and cycling safety. We demonstrated how network constrained spatial analysis methods can be applied to officially reported cycling incident data to identify changes in the spatial-temporal distribution of cycling safety across a transportation network. We also demonstrated how network appropriate spatial analysis techniques can be applied to large, emerging crowdsourced cycling datasets to monitor changes in patterns of ridership. These methods enhance our understanding of the city-wide impact of infrastructure changes on cycling safety and ridership patterns.
# TABLE OF CONTENTS

SUPERVISORY COMMITTEE: ................................................................................................. ii

ABSTRACT.............................................................................................................................. iii

TABLE OF CONTENTS........................................................................................................... v

LIST OF TABLES.................................................................................................................... viii

LIST OF FIGURES................................................................................................................ ix

ACKNOWLEDGEMENTS......................................................................................................... xi

1.0 INTRODUCTION .......................................................................................................... 1
  1.1 Research Context........................................................................................................ 1
  1.2 Research Gap............................................................................................................ 2
  1.3 Research goals and objectives.................................................................................... 4
      References..................................................................................................................... 6

2.0 MONITORING CITY-WIDE PATTERNS OF CYCLING SAFETY ......................... 9
  2.1 Abstract ...................................................................................................................... 9
  2.2 Acknowledgements.................................................................................................... 10
  2.3 Introduction............................................................................................................... 10
  2.4 Methods..................................................................................................................... 12
      2.4.1 Study area......................................................................................................... 12
3.0 USING CROWDSOURCED DATA TO MONITOR CHANGES IN BICYCLE RIDERSHIP: A SPATIAL ANALYSIS

3.1 Abstract

3.2 Acknowledgements

3.3 Introduction

3.4 Study Area and Data

3.4.1 Study Area

3.4.2 Crowdsourced Cycling Data

3.5 Methods
LIST OF TABLES

Table 2.1 Cycling incidents reported to ICBC for Vancouver, Canada from 2009 through 2013

25
LIST OF FIGURES

Figure 2.1 The case study area of Vancouver, BC, Canada. Cycle tracks constructed in 2010 are represented by the green and pink lines. ............................................................ 26

Figure 2.2 a) Planar KDE measures bandwidth in 2-D Euclidean space and estimates density using the five points within the dashed circle. b) Network KDE measures bandwidth along the network and estimates density using the three points along the heavy black line. ......................................................................................................................... 27

Figure 2.3 (a) Planar and (b) network KDE for insurance-reported cycling incidents in Downtown Vancouver in 2013. ........................................................................................ 28

Figure 2.4 Network KDE of cycling incidents from (a) 2009, (b) 2011, (c) 2012 and (d) 2013................................................................................................................................... 29

Figure 2.5 Comparison of network KDE, (a) 2009 and 2011, (b) 2009 and 2012 and (c) 2009 and 2013. Blue areas represent a decrease in intensity of two or more standard deviations. Red areas represent an increase in intensity of two or more standard deviations. Black ellipses highlight areas of consistent change. ...................................... 31

Figure 3.1 Strava users versus cycling population by age in Ottawa-Gatineau (TRANS Committee, 2011). ....................................................................................................... 51

Figure 3.2 Definitions of spatial weights matrices for segment i: (a) first order lag, equal weighting with wij = 1 for contiguous street segments; (b) second order lag with wij = 1 for first order neighbors, wij = 0.5 for second order neighbors; (c) second order lag with wij = 1 for all neighbors. ............................................................................................................. 53
Figure 3.3 Change in relative volume of ridership, May 2015 to May 2016. ............ 54

Figure 3.4 Network local Moran’s Ii of the difference in normalized ridership between May 2015 and May 2016 based on first order neighbors. Pink segments represent clusters of increased ridership and light blue segments represent clusters of decreased ridership. ................................................................. 55

Figure 3.5 Annotated map of local network autocorrelation highlighting areas of infrastructure change during the study period. ............................................. 56

Figure 3.6 Examples of preliminary annotation of changes in spatial patterns of ridership in Ottawa-Gatineau. (a) Clusters of increased ridership and decreased ridership associated with a new multi-use bridge path. (b) Changes in patterns of ridership associated with the closure of a multi-use tunnel. (c) Shift of cyclists from Rideau Canal western pathway to the eastern pathway associated with temporary closures of the western pathway. (d) Shift of ridership from shared roadway bridges to multiuse path. ................................. 58
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Trisalyn Nelson, for her encouragement, support, patience and enthusiasm over the past two years. I would also like to thank Dr. Meghan Winters for her guidance, insight and the direction provided while preparing the manuscripts. I am also thankful for each member of the BikeMaps.org team, past and present. I am honoured to have worked with such a talented team on such a worthy cause. Finally, I am infinitely grateful to my incredible wife, Angela Boss. Thank you for your love, support, encouragement and belief in me.
1.0 Introduction

1.1 Research Context

Bicycling for transportation is beneficial for cyclists, their communities and the environment. Individuals who cycle for transport are less likely to be obese or overweight, have improved cardiovascular health and are at reduced risk for type 2 diabetes (Hu et al., 2003; Gordon-Larsen et al., 2005; Pucher et al., 2010). Social benefits include increased connection to the city, enhanced vitality of cities and reduced traffic congestion (Gehl, J., 2010). Environmental benefits include reduced greenhouse gas emissions, decreased reliance on fossil fuels and a reduction in noise pollution (Pucher & Buehler, 2008).

Cycling as a form of transportation is underutilized in North America, with just one to two percent of all trips made by bike (Pucher & Buehler, 2008; Teschke et al., 2012). In comparison, the percentage of trips made by bicycle in many European nations, including The Netherlands, Denmark and Germany, ranges from ten to thirty percent (Pucher & Buehler, 2008). This discrepancy in ridership levels, combined with the benefits of increasing mode share, suggest that there is a large potential to increase the number of people cycling for transportation in North America (Pucher & Dijkstra, 2003).

Cyclists and potential cyclists cite concern for personal safety as a significant deterrent to cycling (Winters et al., 2011). Safety concerns are warranted given that injury and fatality rates per bike trip are higher than those for automobile trips in Canada and the United States (Beck et al., 2007). Additionally, injury and fatality rates for cyclists in North America are greater than rates in many European countries (i.e., Netherlands,
Denmark and Germany) with higher ridership levels (Pucher & Buehler, 2008). Cyclists perceive bicycle specific infrastructure as safer than regular streets, favoring bike lanes, cycle tracks and neighborhood bikeways over high traffic volume streets with no bicycle facilities (Broach et al., 2012; Dill, 2009; Hood et al., 2011). Research has shown that infrastructure safety improvements lead to increased bicycle use for transportation (Buehler & Dill, 2015). Additionally, increased bicycle use results in the ‘safety in numbers’ effect; as more people cycle, accident rates decrease (Jacobsen, 2003). The disparity in safety and bicycle use for transport, and differences in common cycling infrastructure between North America and many European countries, has driven research into how infrastructure design in Canada and the United States could increase both safety and ridership levels (Pucher & Buehler, 2008).

Many cities are seeking ways to increase the number of people who cycle for transport by making cycling safer for riders of all ages and abilities. A common approach to increasing safety and ridership is the investment of significant capital resources in cycling infrastructure, with some cities investing in connected networks of cycling infrastructure (Buehler & Dill, 2015). To be accountable to the public and encourage political will for future cycling infrastructure projects, it is essential that cities monitor and report the impact of infrastructure on citizens (Handy et al., 2014).

1.2 Research Gap

Traditional approaches to monitoring the impact of infrastructure on safety and ridership are aspatial. Changes to safety are commonly evaluated by comparing incidents along cycling infrastructure to an unimproved control (Lusk et al., 2011), before-after studies (Dill et al., 2011), or by comparing aggregate statistics for large regions over long
time spans (Pedroso et al., 2016). Evaluating changes in ridership are often limited to measuring the number of cyclists using a segment before and after construction of infrastructure (Goodno et al., 2010) or to aggregate statistics for large areas (Buehler & Pucher, 2012). Due to methodological limitations and a lack of spatially continuous data, few spatially explicit approaches have been applied to monitoring change in patterns of safety and ridership following investment in cycling infrastructure.

A key challenge to monitoring change in patterns of safety and ridership has been a lack of appropriate spatial analysis methods. Cycling for transportation occurs on streets, bike lanes, multiuse paths and cycle tracks. These types of infrastructure comprise a one-dimensional network which occupies a subset of two-dimensional space. Spatial pattern analysis methods, useful for quantifying pattern and pattern change, have been developed for two-dimensional space where distance is measured in Euclidean space. Such methods are not necessarily appropriate for one-dimensional network space where distances are best measured along segments of the network (Miller, 1994). Over the past two decades spatial scientists have extended spatial analysis methods to one-dimensional network space. For example, kernel density estimation (Borruso, 2005) and local indicators of spatial autocorrelation (Yamada & Thill, 2007) have been extended to network space. These methods enable the quantification of patterns on a network, but they have rarely been applied in the analysis of cycling safety and ridership patterns, largely due to a lack of data.

A lack of spatially and temporally dense data has hampered the spatial analysis of patterns of safety and ridership following investment in cycling infrastructure. Cycling incident data are typically sourced from insurance claims, law enforcement reports or
hospital records; however, bicycle collisions are relatively rare events and frequently go unreported (Watson et al., 2015; Wegman et al., 2012). Ridership data are commonly gathered through manual or automated counts, which are expensive, time consuming and have limited spatial and temporal resolution. These gaps in safety and ridership cycling data are being addressed through information gathered with crowdsourced applications. Crowdsourcing allows a self-selected group of citizens to collect and report information about their environment or their interactions with their environment. A newly developed web application, BikeMaps.org, provides a central location for citizen to self-report their cycling collisions and near misses (Nelson et al, 2015). Advances in Global Positioning System technology and its incorporation in to mobile devices, such as smartphones, provides a novel source of spatially and temporally dense ridership information. These data are becoming available for analysis (Fertser et al., 2017; Griffin & Jiao, 2015; Heesch & Langdon, 2016; Jestico et al., 2017) and methods for monitoring change in patterns of cycling safety and ridership on a network using these novel sources of data need to be developed.

1.3 Research goals and objectives

The goal of our research was to demonstrate spatially explicit methods for monitoring city-wide changes in patterns of safety and ridership following changes to cycling infrastructure. To meet our goal, we applied network appropriate spatial analysis techniques to a cycling incident dataset for Vancouver, Canada and to a large crowdsourced ridership dataset for Ottawa-Gatineau, Canada.

The first objective (Chapter 2) was to develop a method for monitoring statistically significant changes in patterns of cycling safety following changes in
infrastructure. To meet this objective we first compared planar versus network constrained kernel density estimation for quantifying cycling incident intensity across the study region. Second, we used official reports of cycling incidents as input for network constrained KDE to quantify incident intensity annually from 2009 through 2013. Finally, the resulting network constrained density maps were compared to identify areas of statistically significant change in patterns of cycling safety in the context of newly installed cycle tracks in downtown Vancouver in 2010.

The second objective (Chapter 3) was to develop a method for monitoring statistically significant change in patterns of ridership following changes to cycling infrastructure using a large crowdsourced ridership dataset. To meet our second objective, we demonstrated a need for a spatially explicit approach for identifying city-wide change in the spatial-temporal distribution of cyclists. We quantified network appropriate measures of spatial autocorrelation using three spatial weights matrices and crowdsourced ridership data for Ottawa-Gatineau, Canada. We then calculated and mapped network constrained local indicators of spatial autocorrelation. Finally, we annotated a map of change in patterns of ridership with changes that occurred in the cycling network of Ottawa-Gatineau between May 2015 and 2016 to contextualize how the approach can be used to monitor city-wide changes in patterns of ridership.
References


2.0 MONITORING CITY-WIDE PATTERNS OF CYCLING SAFETY

2.1 Abstract

Many cities are making significant financial investments in cycling infrastructure with the aim of making cycling safer for riders of all ages and abilities. Methods for evaluating cycling safety tend to summarize average change for a city or emphasize change on a single road segment. Few spatially explicit approaches are available to evaluate how patterns of safety change throughout a city because of cycling infrastructure investments or other changes. Our goal is to demonstrate a method for monitoring changes in the spatial-temporal distribution of cycling incidents across a city. Using cycling incident data provided by the Insurance Corporation of British Columbia, we first compare planar versus network constrained kernel density estimation for visualizing incident intensity across the street network of Vancouver, Canada. Second, we apply a change detection algorithm explicitly designed for detecting statistically significant change in kernel density estimates. The utility of network kernel density change detection is demonstrated through the comparison of cycling incident densities following the construction of two cycle tracks in the downtown core of Vancouver. The methods developed and demonstrated for this study provide city planners and transportation engineers a means of monitoring city-wide change in the intensity of cycling incidents following enhancements to cycling infrastructure or other significant changes to the transportation network.
2.2 Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

2.3 Introduction

Many cities are seeking ways to increase the number of people who cycle for transport by making cycling safer for riders of all ages and abilities. While cycling has numerous physical, environmental and social benefits, ridership levels remain low in North America (Gordon-Larsen et al., 2005; Pucher & Buehler, 2008; Teschke et al., 2012). In Canada and the United States, approximately one to two percent of all trips are taken by bike (Pucher & Buehler, 2008; Teschke et al., 2012). Cyclists and potential cyclists frequently cite concern for personal safety as a significant deterrent to bicycling (Winters et al., 2011a). Research has shown that infrastructure safety improvements, such as the installation of bike lanes, bike specific pathways and cycle tracks, lead to increased bicycle use for transportation (Buehler & Dill, 2015). Additionally, increased bicycle use results in the ‘safety in numbers’ effect; as more people cycle, incident rates decrease (Jacobsen, 2003).

To overcome barriers to increased ridership, cities are making significant investments in cycling infrastructure, with many cities making investments in connected networks of infrastructure. To be accountable to the public and encourage political will for cycling infrastructure projects, it is essential that cities monitor and report the impact of infrastructure on citizens. Safety impacts can bring both health and economic benefits (Krizee, 2007; Mueller et al., 2015). Standard approaches to monitoring safety quantify
change in incidents for an entire city (Pedroso et al., 2016) or on a single street segment or intersection (Chen, et al., 2012; Dill et al., 2011). However, approaches to characterize changes to safety across a city’s transportation network are limited. Mapping change in safety across the network can show where increases and decreases in incidents are occurring, and account for shifts from one street to the next as cyclists alter routes to use bicycling infrastructure.

A challenge in evaluating network level changes in cycling safety is that cycling collisions are mapped as point locations. Most of the methods designed for analysis of point data are unsuitable for phenomena constrained to network space (Yamada & Thill, 2004). Over the past two decades, spatial scientists have extended many traditional point pattern analysis methods to one dimensional, network space. For example, Okabe and Yamada (2001) developed a network specific K-function. Several others have developed kernel density estimation (KDE) techniques suitable for network based analysis which use network distances instead of Euclidean distances (Borruso, 2005; Okabe et al., 2009).

Network constrained KDE has been applied in a variety of disciplines including analysis of economic activities (Produit et al., 2010), traffic incidents (Harirforoush & Bellalite, 2016; Xie & Yan, 2008) and cycling infrastructure planning (Lachance-Bernard et al., 2011). Network KDE has also been integrated with local measures of spatial autocorrelation in the analysis of traffic incidents (Xie & Yan, 2013). While standard KDE has been used as a visualization tool for cycling incident density (Delmelle & Thill, 2008), the application of network KDE to studies of cycling incidents is very limited. Network KDE has potential to advance spatially explicit methods of monitoring change in cycling incidents throughout a city, which is beneficial when evaluating the impact of
cycling infrastructure. Assessment of infrastructure enhancements is often limited to a small set of road segments where comparisons are made using space for time substitutions. Few spatially explicit approaches exist for evaluating changes in the distribution of cycling incidents across a city following improvements to cycling infrastructure.

Our goal is to develop a method for monitoring statistically significant changes in the spatial and temporal variation of cycling incidents following changes in infrastructure. We analyzed cycling incident data from the city of Vancouver, Canada, from January 1, 2009 to December 31, 2013 according to the following objectives. First, we compared the suitability of planar KDE versus network constrained KDE for measuring cycling incident intensity across the study region. Second, official reports of cycling incidents were used to quantify incident intensity annually from 2009 through 2013. Third, the resulting network constrained density maps were compared to identify areas with statistically significant change in the intensity of cycling incidents following the installation of cycle tracks in downtown Vancouver in 2010.

2.4 Methods

2.4.1 Study area

The case study area is the city of Vancouver, Canada, with a population of 603,000 (Statistics Canada, 2011a) (Figure 2.1). Vancouver’s mild climate is favorable to cycle commuting year round and 4.4% of workers commute by bicycle (Statistics Canada, 2011b). Monthly average minimum temperatures are greater than 0 °C in the winter and monthly average maximum temperatures below 23 °C in the summer, though
the city receives a significant amount of precipitation, averaging nearly 1200mm per year (Government of Canada, 2010).

2.4.2 Transportation infrastructure

The city has a wide variety of transportation infrastructure including arterial, collector and local streets. Vancouver has been promoting cycling as a safe and convenient mode of transportation since 1988 (Vancouver, 1988). Historically, Vancouver’s primary emphasis has been the development of local street bikeways (Vancouver, 1999), but the downtown core, which is our region of focus, has few local streets that are used as bikeways. In 2009, there were mainly painted bike lanes downtown. In 2010 dedicated cycle tracks were installed along two major corridors (Figure 2.1). Since the time frame of this case study there has been substantial investment in a cycling network downtown; however, we could not extend our study period since ICBC cycling incident data beyond 2013 was not available.

Transportation infrastructure data was obtained from the city of Vancouver’s Open Data catalogue (Vancouver, 2017). The portion of the network used in this study consists of 1763 street segments and 1119 nodes. The network data was preprocessed to ensure correct topology and subsequently modeled as an undirected graph. Cycling network data was also obtained from the city’s Open Data catalogue (Vancouver, 2017).

2.4.3 Cycling incident data

The cycling incident data were sourced from the Insurance Corporation of British Columbia (ICBC), the provincial insurance provider supplying mandatory coverage to all motor vehicles in BC. The data contains all reported crashes between bicycles and motor
vehicles from 2009 to 2013 (Table 2.1). The location of incidents are reported as street addresses or intersections which were geocoded to the street network.

2.4.4 Kernel density estimation

In order to detect change in cycling safety, we first mapped spatial variation in cycling safety along a network in two time periods \( t_0 \) and \( t_1 \) and then quantified change between \( t_0 \) and \( t_1 \). The standard implementation of KDE is used to produce a smoothed density surface from point events in two-dimensional space. A grid surface is superimposed on a study area. A kernel function is used to calculate the density of point events for the centroid of each cell in the grid. The kernel function weights points within a circle of influence according to the Euclidean distance between the centroid and the points. The general form of a kernel estimator is:

\[
\lambda(s) = \sum_{i=1}^{n} \frac{1}{r^2} k\left(\frac{d_{is}}{r}\right)
\]

where \( \lambda(s) \) is the density at the location of measurement \( s \), \( r \) is the bandwidth or smoothing parameter, \( d_{is} \) is the distance between \( s \) and point \( i \), \( k \) represents a kernel function that weights the value of \( i \) at \( s \) and \( n \) is the number of events within the bandwidth from location \( s \). A variety of kernel functions are commonly used including Gaussian, Quartic, and Epanichnekov. The choice of kernel function has less impact on the final density surface than the choice of bandwidth (O'Sullivan & Unwin, 2002; Silverman, 1986).

The challenge with the standard implementation of the KDE for cycling safety is that mapping to planar space tends to be good for identifying hot spots of safety concern.
at intersections, but linear features along road corridors can be missed due to the circular geometry of bandwidths. Constraining a kernel density estimator by a network uses distances along the network as opposed to Euclidean distance for creating the bandwidth (Figure 2.2).

The general form of a network constrained kernel estimator is:

\[ \lambda(s) = \sum_{i=1}^{n} \frac{1}{r} k \left( \frac{d_{is}}{r} \right) \]

where \( r \) and \( d_{is} \) are measured over the network. The resulting intensity value is based on linear units instead of areal units. As for standard KDE, the choice of kernel function \( k \) for network KDE is less important to the resulting density estimate than the choice of bandwidth \( r \) (Xie & Yan, 2008).

Two primary methods have been developed for network constrained KDE, differing in their choice of the basic spatial unit (BSU) of analysis. One method overlays a grid on the study area and uses a grid cell as the BSU (Borruso, 2008; Produit et al., 2010). The second method attempts to divide the network into segments of equal length and uses these as the BSU (Xie & Yan, 2008). The division of a network into basic spatial units of equal length is non-trivial and results in residual segments which are shorter than the defined lixel size (Xie & Yan, 2008). Furthermore, as the underlying network changes, such as through the construction of new streets or cycling paths, the absolute position of lixels within the network may not be consistent between two time periods. We elected to use the grid cell as our basic spatial unit as it has the advantage of producing a surface where the location of grid cells is invariant with respect to changes in
the underlying network over time. This facilitates the spatial-temporal comparison of the network KDEs.

2.4.5 Comparison of Planar KDE and Network KDE

To demonstrate the difference in bicycle safety patterns mapped with standard and network constrained KDEs, density surfaces were created using both the standard and network KDE methods applied to cycling incidents from 2013 in the downtown core of Vancouver, Canada. Density surfaces were generated with a grid cell size of 50 meters and a bandwidth of 300 meters. Bandwidth strongly influences the resulting density surface. A bandwidth that is too small results in an undersmoothed surface and too large a bandwidth will oversmooth the surface. We initially selected a bandwidth of 450m based on Silverman’s rule of thumb (Silverman, 1986) which estimates a smoothing parameter that minimizes mean integrated squared error assuming the underlying density is Gaussian. Silverman’s rule of thumb has a tendency to overestimate the bandwidth when the underlying distribution is not Gaussian; as such, we reduced the bandwidth to 300m. We performed a visual comparison of the resulting density surfaces, examining the location, shape, size and intensity of hot spots, to assess the utility of both methods for visualizing cycling incident density.

2.4.6 Network KDE change detection

To detect change in cycling safety through time, density surfaces were created using network KDE applied to the cycling incident data for each year including 2009, 2011, 2012 and 2013. The data from 2010 was not included as the cycle tracks in the downtown core of Vancouver were under construction during this year. We generated
density surfaces using a grid cell size of 50m. To enable comparison of density surfaces, we used a common smoothing parameter for all surfaces (Bowman & Azzalini, 1997).

After generating density surfaces for each year, the surfaces from 2011 through 2013 were compared to 2009. We compared the difference of the square root density estimates at a location divided by the standard error (Bowman & Azzalini, 1997). Specifically, the change in grid cell $i$ over time is calculated as:

$$\text{change}_{\Delta t} = \frac{\sqrt{\lambda_{i,t_1}} - \sqrt{\lambda_{i,t_0}}}{\sqrt{se_{t_1}^2 + se_{t_0}^2}}$$

where $\lambda_{i,t_0}$ and $\lambda_{i,t_1}$ are the kernel density estimates of cycling incidents for each grid cell $i$, in time periods $t_0$ and $t_1$, and $se_{t_0}$ and $se_{t_1}$ are the standard errors of the kernel density estimates for time periods $t_0$ and $t_1$.

The use of the variance stabilizing square root transformation and standard error allows the difference between KDEs to be expressed in terms of standard deviation (SD) (Bowman & Azzalini, 1997). Areas of the resulting comparison which have SD $<-2$ or SD $>2$ represent locations where cycling incidents have either decreased or increased more than expected given the assumption that the underlying point distributions are equal between times $t_0$ and $t_1$.

2.5 Results

2.5.1 Comparison of Planar KDE and Network KDE

Planar KDE generated circular or ellipsoidal shaped hot spots (regions of high intensity of cycling incidents) (Figure 2.3). In contrast, network KDE emphasized linear
hot spots of cycling incidents. For the selected bandwidth, the regions of highest intensity for planar KDE are much greater in extent than those for network constrained KDE. Additionally, the planar KDE appears to be more heavily smoothed than the network KDE. Furthermore, network KDE identified fewer high intensity regions. The tendency of planar KDE to overestimate density (Xie & Yan, 2008) combined with network KDE emphasizing linear hotspots, suggests that network KDE is preferable for analysis of cycling incidents on a street network.

2.5.2 Network kernel density change detection

The result of the network constrained KDE of cycling incidents from 2009, 2011, 2012 and 2013 are shown in Figure 2.4. The density surfaces displayed spatial variation over the study period years in the areas of maximum intensity. For example, the 2009 density surface (panel a) displayed a region of high cycling incident intensity along Burrard Street, the corridor which is one block to the northwest of the planned location of the Hornby Street cycle track. Following construction of the Hornby Street cycle track in 2010, the cycling incident intensity along the Burrard Street corridor was not present in the maps for 2011, 2012, and 2013 (panels b-d), although increased incident intensity was present at the southern end, especially in 2011 and 2012). Locations of low cycling incident intensity remained consistent across the study period.

Beyond visual inspection for change over time, we also looked at whether change was statistically significant, comparing each of the network KDEs from 2011 through 2013 to 2009. The result of each comparison was a map displaying change in terms of standard deviation (Figure 2.5). Near the cycle tracks installed in 2010, there were several
regions where changes were statistically significant and consistent over time. Region 1 in Figure 2.5, located at the northern end of the Hornby Street cycle track, displayed an increase in cycling incident intensity (greater than two standard deviations) in 2011, 2012 and 2013 as compared to 2009. Similarly, region 2 highlighted an area north of the Dunsmuir Avenue cycle track where cycling incident intensity significantly increased. Conversely, region 3 nearby the Hornby Street cycle track denotes an area where incident intensity decreased in two out of three years following cycle track construction.

Additionally, Figure 2.5 indicated several areas of consistent change that were not in the immediate vicinity of the cycle tracks (regions 4-6). Regions 4 and 5 denoted areas of consistent increase in intensity in 2011 through 2013 as compared to 2009, while region 6 highlighted an area with decreased in 2011 and 2012 as compared to 2009.

2.6 Discussion

We demonstrated a method for monitoring change in the spatial-temporal distribution of cycling incidents across a large geographic area, and applied this to a case study of the installation of cycle tracks in downtown Vancouver, Canada. Following the installation of cycle tracks, we found areas of statistically significant increase and decrease in cycling incident intensity near the new infrastructure. We also found areas of increased and decreased intensity in areas not associated with the new cycle tracks. The methods we describe are intended to be exploratory in nature; they identify areas of statistically significant change in cycling safety, but do not explain reasons for the changes.
Results suggest mapping cycling safety incidents with a linear search function using network distances is preferable to a circular search function that uses Euclidean distances, in that incorporating the network enables identification of linear hot spots. This result is consistent with the findings of Borruso (2008) who implemented a network KDE method using a uniform kernel, applied to bank and insurance company locations to identify clusters of financial services and test for the presence of a financial district. Similarly, Xie and Yan (2008) demonstrated that their implementation of network KDE, using a distance weighted kernel function, was more appropriate for estimating the density of traffic incidents in network space than planar KDE. In addition to emphasizing linear corridors, network KDE is less likely to overestimate intensity than planar KDE (Xie & Yan, 2008). Respect for the one-dimensional nature of cycling infrastructure, emphasis on linear corridors, and a lack of overestimating intensity suggest that network KDE should be used by city planners, transportation engineers and researchers when estimating cycling incident intensity instead of planar KDE.

Our results indicate that for a given bandwidth, a planar KDE will be over smoothed, relative to the network KDE. Over smoothing is due to the circular form of the kernel function and measurement of distance in Euclidean space which results in a larger number of cycling incidents contributing to the intensity for any given grid cell as compared to the linear search function and network distances employed by network KDE. The larger bandwidth includes more events and increases the smoothness of the resulting density surface. In contrast, a smaller bandwidth limits the number of points included leading to a less smoothed surface and emphasis on more localized patterns.
(O’Sullivan & Unwin, 2002). The substantive impact of bandwidth raises questions regarding methods for estimating the optimal bandwidth for network constrained KDE.

A variety of techniques have been developed for estimating the optimal bandwidth for planar KDE including rules of thumb, cross-validation techniques and plug-in methods (Sheather, 2004), but these methods may not be suitable for network KDE (Okabe et al., 2009). We initially selected a bandwidth of 450m for our analysis based on Silverman’s rule of thumb (Silverman, 1986). It is known that this method has a tendency to estimate too large a bandwidth when the underlying point pattern is non-Gaussian and our initial planar KDE were oversmoothed (Sheather, 2004). We ultimately used a bandwidth of 300m to balance oversmoothing planar KDE versus undersmoothing network KDE. It is clear that any chosen bandwidth is unlikely to be optimal for both planar and network KDE. Little, if any work has been published on methods of selecting optimal bandwidths for studying point patterns on networks, thus further study of this issue is warranted.

We identified several areas of statistically significant increase and decrease in safety near the newly installed cycle tracks. The decrease in cycling incident intensity on Burrard Street following the installation of the Hornby Street cycle track one block east is not unexpected. Previous research indicates that cyclists will detour about 10% from a shortest path route in order to make use of cycling specific infrastructure (Winters et al., 2011b). Additionally, studies have shown that the risk of injury on a cycle track is far lower than in the street (Lusk, et al., 2011; Teschke et al., 2012). As such, it is reasonable to expect some cyclists to alter their route from Burrard Street to the cycle track one block east, which could result in a decrease in incident density as bicyclists benefit from
safer infrastructure. Similar logic may apply to other areas where we identified increases in incident intensity. Additionally, the construction of new infrastructure may be promoting an increase in the number of people cycling for transportation (Pedroso et al., 2016). The increased density of cycling incidents in these areas may not represent an increase in risk. Rather, the change may simply be due to increased exposure resulting from higher volumes of cyclists.

This discussion regarding changes in the density of cycling incidents highlights an important limitation of the methods we demonstrated; the KDE do not incorporate ridership volumes. This is not unusual in investigations of cycling incidents as detailed spatial-temporal data on cyclist volumes are rarely available (Loidl et al., 2016). Furthermore, the methods described herein are not intended to be explanatory in nature. Rather, we present this work as a tool for exploring changes in the spatial-temporal pattern of cycling incidents to guide subsequent investigation into the nature or cause of these changes. If exposure data were available, such as the volume of cyclists per road segment across the study region, it could be incorporated into the network constrained KDE by applying a uniform network transformation using the exposure data to weight the length of each segment and location of point events (Okabe & Satoh, 2006). The network with transformed data can then be used as input for analysis. The ability to easily incorporate exposure data is another compelling reason for using network KDE to study cycling incident intensity as compared to planar KDE.

Our goal was to develop a method for monitoring statistically significant changes in the spatial and temporal variation of cycling incidents following investment in cycling infrastructure, but the methods are not restricted to change detection following
infrastructure enhancement. The methods presented here can be used more generally to identify changes in cycling incident density over time. Our change detection maps identified several areas of consistent change that may not be related to the installation of cycling infrastructure. It is possible that ridership volumes have either increased or decreased in these areas. As multiple years of data are being compared to a single year, it is also possible that random variation in the density of cycling incidents accounts for the observed changes. Underreporting of bicycle collisions in combination with most reporting systems only incorporating incidents with vehicles leads to sparse data (Watson et al., 2015; Wegman et al., 2012). As such, it is possible for a random increase or decrease of events near a single location to obscure the true underlying density. The scarcity of data may be addressed in several ways. For example, if studying change before and after a specific point in time, one may incorporate multiple years of data, if available, to represent times $t_0$ and $t_1$. Another possibility that may be available in the future is the incorporation of non-traditional data sources, such as data from crowdsourcing initiatives, which collect both collision and near miss data (Nelson et al., 2015).

### 2.7 Conclusions

Cycling is a sustainable mode of transportation with health, environmental and social benefits. In order to make cycling a viable transportation option for people of all ages and abilities, many cities are attempting to increase safety through financial investment in cycling infrastructure. While installation of bicycle specific infrastructure may increase safety, few spatially explicit approaches exist for assessing how patterns of safety change across a city following such improvements. We demonstrated a method,
network KDE change detection, that provides city planners, transportation engineers and researchers a means of monitoring change in cycling incident intensity following enhancements to cycling infrastructure or other significant changes to the transportation network. The methods demonstrated in this paper are well suited for studies that include exposure data and will become more robust as datasets on cycling safety grow.
Table 2.1 Cycling incidents reported to ICBC for Vancouver, Canada from 2009 through 2013.

<table>
<thead>
<tr>
<th>Year</th>
<th>Cycling Incidents in Study Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>176</td>
</tr>
<tr>
<td>2011</td>
<td>210</td>
</tr>
<tr>
<td>2012</td>
<td>200</td>
</tr>
<tr>
<td>2013</td>
<td>205</td>
</tr>
</tbody>
</table>
Figure 2.1 The case study area of Vancouver, BC, Canada. Cycle tracks constructed in 2010 are represented by the green and pink lines.
Figure 2.2 a) Planar KDE measures bandwidth in 2-D Euclidean space and estimates density using the five points within the dashed circle. b) Network KDE measures bandwidth along the network and estimates density using the three points along the heavy black line.
Figure 2.3 (a) Planar and (b) network KDE for insurance-reported cycling incidents in Downtown Vancouver in 2013.
Figure 2.4 Network KDE of cycling incidents from (a) 2009, (b) 2011, (c) 2012 and (d) 2013.
Figure 2.5 Comparison of network KDE, (a) 2009 and 2011, (b) 2009 and 2012 and (c) 2009 and 2013. Blue areas represent a decrease in intensity of two or more standard deviations. Red areas represent an increase in intensity of two or more standard deviations. Black ellipses highlight areas of consistent change.
References


location applied to Ljubljana. Lecture *Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6783 LNCS(PART 2), 136–150. doi: 10.1007/978-3-642-21887-3_11


3.0 USING CROWDSOURCED DATA TO MONITOR CHANGES IN
BICYCLE RIDERSHIP: A SPATIAL ANALYSIS

3.1 Abstract

Cycling is a sustainable mode of transportation with numerous health, environmental and social benefits. Cities are increasing investments in cycling specific infrastructure with the goal of increasing ridership. New infrastructure has the potential to impact the upgraded corridor, as well as nearby street segments and cyclist flow across the city. Evaluation of the impact of new infrastructure is often limited to manual or automated counts of cyclists before and after construction, or to aggregate statistics for a large region. Due to methodological limitations and a lack of spatially continuous data, few spatially explicit approaches have been applied to evaluate how patterns of ridership change following investment in cycling infrastructure. Our goal is to demonstrate the application of spatial analysis methods to emerging sources of crowdsourced cycling data to monitor changes in the spatial-temporal distribution of cyclists across a city. Specifically, we use crowdsourced ridership data from Strava to examine changes in the spatial-temporal distribution of cyclists in Ottawa-Gatineau, Canada, using network-constrained local indicators of spatial autocorrelation. Our results indicate that the installation or removal of cycling infrastructure in one location affects the flow and relative volume of cyclists at multiple locations and that these changes are evident in crowdsourced ridership data. City planners, transportation engineers and researchers can use the methods outlined here to monitor city-wide changes in ridership patterns following investment in cycling infrastructure or other changes to the transportation network.
3.2 Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada. We would like to thank the City of Ottawa for their assistance with data collection and valuable insights.

3.3 Introduction

Cycling is a sustainable mode of transportation with numerous health, environmental and social benefits (Gordon-Larsen et al., 2005; Pucher & Buehler, 2008; Teschke et al., 2012). Despite benefits, ridership levels in North America remain low, with just 1-2% of commutes made by bicycle (Pucher & Buehler, 2008). Studies have shown that cyclists and potential cyclists favor bicycle specific infrastructure that is segregated from traffic or adjacent to low volume and low speed traffic (Broach et al., 2012; Winters et al., 2011). In an effort to increase ridership, many cities are making significant financial investments in cycling infrastructure, and several cities are developing cycling infrastructure networks (Buehler & Dill, 2015). It is essential that cities monitor and report on the impact of infrastructure projects on ridership to be accountable to the public and to encourage political will for future investments in cycling infrastructure (Handy et al., 2014).

Monitoring and evaluation of the impacts of investment in cycling infrastructure across a city have been difficult due to a lack of spatially explicit ridership data. Traditional sources of cycling data and route information include manual counts, intercept surveys, automated pneumatic tube counters, mail-back surveys and recruitment into studies (Forsyth et al., 2010; Hyde-Wright et al., 2014; Nordback et al., 2013), but
these sources lack the spatial and temporal density required for detailed spatial analysis. Data limitations are being overcome through advances in Global Positioning System (GPS) technology and its incorporation into portable devices, such as smartphones, which provide a novel source of spatially and temporally dense cycling data. Further, large citizen science cycling datasets are becoming available for analysis. For example, in North American cities, crowdsourced cycling data has been used to examine where cycling for health occurs with respect to land use diversity, bicycle facilities and residential and employment density (Griffin & Jiao, 2015) and researchers have observed a strong correlation between crowdsourced fitness app data and manual cycling counts (Jestico et al., 2016).

Methodological limitations have also hampered city-wide evaluation of changes in patterns of ridership. Cycling occurs on transportation networks which occupy a subset of two-dimensional space. Traditional spatial pattern analysis methods, useful for quantifying pattern and pattern change, have been developed for two-dimensional space and are not necessarily appropriate for one-dimensional network space (Miller, 1994). Analytical limitations are being addressed as spatial scientists extend traditional spatial pattern analysis methods to network space. For example, Borruso (2005) extended kernel density estimation to networks, Okabe and Yamada (2001) introduced a network specific K-Function, and Yamada and Thill (2007) applied local indicators of spatial autocorrelation to network space.

Our goal is to demonstrate how network-based spatial pattern methods can be applied to crowdsourced ridership data to monitor changes in the spatial-temporal variation of ridership across a city. We analyzed a large crowdsourced cycling dataset for
Ottawa-Gatineau, Canada, comparing volumes of cyclists from May 2015 and May 2016 to meet the following objectives. First, we demonstrated the need for a spatially explicit approach for identifying statistically significant change in the spatial-temporal variation of cyclists across the city. Second, we quantified change in patterns of ridership using network appropriate measures of global and local spatial autocorrelation. Third, working with stakeholders, we annotated maps of change in the spatial patterns of ridership with changes that occurred in the cycling network between May 2015 and 2016, and used this case study to illustrate how this approach can be used to monitor spatial variation in ridership.

3.4 Study Area and Data

3.4.1 Study Area

The case study area is Ottawa-Gatineau, Canada with a population of 1.24 million (Statistics Canada, 2011a). Approximately 2.2% of workers commute by bicycle (Statistics Canada, 2011b). The region has invested significant financial resources in bicycle and multi-use infrastructure over the past several years and currently has over 600 km of bicycle paths (National Capital Commission, 2017).

3.4.2 Crowdsourced Cycling Data

The City of Ottawa has partnered with Strava, a social network for runners and cyclists, to obtain a large crowdsourced cycling dataset. The Strava mobile App is used by athletes to track their activities which are then uploaded to the Strava website. This volunteer sourced data is anonymized and aggregated into the Strava Metro data product (Strava Metro, 2017). The Strava Metro data used in this study consists of activity counts
(bicycle trips) per segment of transportation infrastructure in the Ottawa-Gatineau region, aggregated for weekdays in May 2015 and May 2016. We chose this time period as several substantial changes were made to cycling infrastructure between May 2015 and 2016, and thus this serves as a pre-post analysis. There were a total of 4.49 million activity counts from 52,123 bike trips across 71,205 network segments. Strava is used by both commuters and recreational cyclists. Typically, about fifty percent of activities in dense, metropolitan areas are commutes (Strava Metro, 2017). In Ottawa-Gatineau we expect a higher proportion of commuters due to a marketing campaign led by the city for commuters to contribute data to Strava in advance of the study period. The street segment map included in the Strava Metro data product was derived from OpenStreetMap (OpenStreetMap, 2017).

There are differences between Strava users and the general cycling population by gender and age. The percentage of male Strava users (78.2%) is higher than the percentage of male cyclists in the Ottawa-Gatineau region (68%). Strava users in the 25-34 and 35-44 age groupings are over-represented as compared to the actual cycling population, while the under 25, 55-64, 65-74 and over 75 age groupings of Strava users are under-represented (Figure 3.1).

3.5 Methods

3.5.1 Changes in ridership

To map change in ridership along each segment we first summed the total of all activity counts across the study area for each time period, and then calculated a normalized ridership value for each segment, representing the proportion of all activity
counts that occurred within that time period on each segment. We subtracted the normalized ridership in May 2015 from May 2016 on a segment-by-segment basis and created a map of the absolute difference. We visualized the resulting data on a map in an attempt to identify change in the spatial variation of bicycle trips.

3.5.2 Clustering of changes in ridership across the study area

We used a spatial statistic that measures autocorrelation to identify where changes in ridership occurred that were unexpected based on a null hypothesis of random change. Spatial autocorrelation is the concept that all things are related, but things near to one another are more related than things far apart (Tobler, 1965). Positive spatial autocorrelation, often described as a cluster, is present when the value of a variable at a location is similar to values of the same variable at locations close by. Negative spatial autocorrelation, often described as an outlier, is present when the values of a variable at nearby locations are dissimilar.

Moran’s $I$ is a common measure of spatial autocorrelation and is useful for identifying spatial autocorrelation in values that are high or low relative to the mean. By focusing on high and low values, Moran’s $I$ is helpful for identifying if large or small changes in the relative volume of cyclists is positively or negatively autocorrelated. If, for example, change in ridership is not spatially autocorrelated changes are likely resulting from random patterns; however, clustered change in ridership is likely the result of non-random change. The Moran’s $I$ statistic is defined as:

$$I = \frac{n \sum_i^n \sum_j^n w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_i^n \sum_j^n w_{ij}) \sum_i^n (x_i - \bar{x})^2},$$
where \( n \) is the number of regions, \( x_i \) is the variable value at region \( i \) for \( i = (1, \ldots, n) \), \( x_j \) is the variable value at neighboring region \( j \) for \( j = (1, \ldots, n) \), \( \bar{x} \) is the average of variable \( x \) across the study area, and \( w_{ij} \) is the \( i^{th} - j^{th} \) element of a weight matrix \( W \) designating the spatial relationship between regions \( i \) and \( j \) (Moran, 1948). The spatial relationship between two regions can be defined in several different ways. One option is a binary adjacency matrix where \( w_{ij} = 1 \) for regions that are adjacent, otherwise \( w_{ij} = 0 \). To expand the size of the neighborhood of \( i \), the neighbors of \( j \) can be added to the neighborhood of \( i \). This process can be iterated to generate multiple neighborhood sizes (lags). An alternative is to define the relationship in terms of proximity where the value for \( w_{ij} \) is weighted based on the distance between regions \( i \) and \( j \). In general, the choice of spatial weights matrix should reflect your conceptualization of the interaction of regions or features in space.

Generally, Moran’s \( I \) is implemented on a two-dimensional surface; however, our study area consists of transportation infrastructure which occupies one-dimensional network space. As such, we based our spatial weights matrix on the contiguity of infrastructure segments. Every segment has two nodes that represent either an intersection or the end of a segment. Therefore, the first order lag of segment \( i \) includes those segments \( j \) which share a node with \( i \). To extend the neighborhood to a second order lag, the neighbors of each segment \( j \) are also included as neighbors of \( i \).

All spatial methods are sensitive to definitions of what is nearby. As “nearby” gets larger the scale of the pattern being quantified becomes coarser. Determining the appropriate definition of nearby can be somewhat subjective, therefore we implemented three definitions of neighborhood to calculate a network Global Moran’s \( I \) (Figure 3.2):
1) First order lag with $w_{ij} = 1$ for contiguous street segments and $w_{ij} = 0$ otherwise;

2) Second order lag where $w_{ij} = 1$ for first order neighbors, $w_{ij} = 0.5$ for second order neighbors and $w_{ij} = 0$ otherwise;

3) Second order lag with $w_{ij} = 1$ for first and second order neighbors and $w_{ij} = 0$ otherwise.

Once the spatial weights matrices have been established, the calculation of the Moran’s $I$ statistic for network space is identical to that for planar space. We used the difference in normalized ridership between May 2015 and 2016 as input for calculating the global Moran’s $I$ statistic.

### 3.5.3 Identifying specific areas with changes in ridership

While global Moran’s $I$ quantifies spatial autocorrelation for an entire study area with a summary statistic, local Moran’s $I_i$ quantifies spatial autocorrelation for individual spatial units in the study region allowing identification of local spatial patterns or clusters. Local Moran’s $I_i$ is calculated as:

$$I_i = \frac{n(x_i - \bar{x}) \sum_j^n w_{ij}(x_j - \bar{x})}{\sum_i^n (x_i - \bar{x})^2},$$

where $n$ is the number of regions, $x_i$ is the variable value at region $i$ for $i = (1, \ldots, n)$, $x_j$ is the variable value at neighboring region $j$ for $j = (1, \ldots, n)$, $\bar{x}$ is the average of variable $x$ across the study area, and $w_{ij}$ is the $i^{th} - j^{th}$ element of a weight matrix $W$ designating the spatial relationship between regions $i$ and $j$ (Anselin, 1995). Local Moran’s $I_i$ measures of spatial autocorrelation is particularly useful for identifying clusters (positive spatial autocorrelation) and outliers (negative spatial autocorrelation) in high and low values. As
such, Local Moran’s $I_i$ can be used to find clusters of increased or decreased ridership, and outliers or segments of the network that have a different pattern of ridership relative to what surrounds it. We calculated local Moran’s $I_i$ using a first order lag neighborhood and the difference in normalized ridership between May 2015 and 2016. We chose a first order lag neighborhood based on the results of global Moran’s $I$.

3.5.4 Relating changes in ridership to changes in infrastructure

To contextualize if and how network based measures of local spatial autocorrelation applied to crowdsourced cycling data can be used to monitor ridership, we worked with municipal staff from Ottawa, Canada, to annotate a map of change with knowledge of locations of cycling infrastructure construction and closures in the Ottawa-Gatineau region between May 2015 and 2016. Major infrastructure changes pertaining to cycling during this time period include the installation of multiuse pedestrian and cycling bridges, new off street bike paths, and re-routing of existing infrastructure and temporary closures of cycling infrastructure due to both the construction of light rapid transit and reconstruction of existing infrastructure.

3.6 Results

3.6.1 Changes in ridership

The map of the absolute difference in normalized ridership between May 2015 and May 2016 is presented in Figure 3.3. We were able to identify network segments where normalized ridership increased or decreased, but it was not possible to determine if the observed changes were due to chance or to identify where statistically significant network clustering was present in the variation in ridership.
3.6.2 Clustering in changes in ridership across the study area

Global Moran’s $I$ indicated that statistically significant network clustering was present in the normalized change of ridership between May 2015 and 2016. The Moran’s $I$ statistic was higher for a spatial neighborhood defined by first order lags ($I = 0.593$, $p = <0.001$), as compared to weighted ($I = 0.492$, $p = <0.001$) and unweighted ($I = 0.456$, $p = <0.001$) second order lags. The Moran’s $I$ results suggest clustering is more pronounced when neighborhoods, or the definition of “nearby”, were small.

3.6.3 Identifying specific areas with changes in ridership

The map of network local Moran’s $I$, indicated the presence of positive network autocorrelation in the difference of normalized ridership between May 2015 and 2016 (Figure 3.4). Positive network autocorrelation is visualized as clusters of segments where the relative volume of ridership increased, but also as clusters where the relative volume of ridership decreased. The map also indicates several outliers, which are segments that showed an increase or decrease, but neighboring segments had an opposite change in ridership.

3.6.4 Relating changes in ridership to changes in infrastructure

Upon annotating the maps of Local Moran’s $I$, we observed changes in patterns of ridership associated with changes in infrastructure (Figure 3.5). Following the installation of a multiuse pedestrian and cycling bridge (Figure 3.6a), our results indicate an increase in the proportion of ridership on the segments near the new bridge, and a decrease on a bridge shared with motor vehicle traffic 0.7 km to the north. We also observed changes in ridership patterns following the temporary closure of a multi-use
tunnel (Figure 3.6b) which was part of an east-west corridor frequently used by cyclists. Construction and closures at two points along the Rideau Canal western pathway led to a significant decrease in ridership on a large portion of the pathway and a significant increase on the eastern pathway (Figure 3.6c). Between May 2015 and May 2016, a new pedestrian and cycling pathway was installed along the MacDonald-Cartier Bridge between Gatineau and Ottawa. This analysis showed that ridership has shifted to the new multiuse pathway from both the MacDonald-Cartier and Interprovincial/Alexandra bridges (Figure 3.6d). Similar results were observed across the study area following improvements to cycling infrastructure and temporary closures with detours due to construction (Figure 3.5).

3.7 Discussion

The ability of cities to monitor and evaluate the impacts of investment in cycling infrastructure has been hampered by methodological limitations and a lack of spatially continuous ridership data. In this paper, we demonstrated a spatially explicit approach for monitoring variation in ridership by applying network constrained spatial analysis methods to novel sources of cycling data. We applied this in a real-world example, using network autocorrelation analysis on a crowdsourced cycling ridership dataset. In this case study, we found in locations where cycling infrastructure had been installed or removed, we could detect changes in the flow and relative volume of ridership at nearby locations.

We demonstrated a spatially explicit approach for monitoring variation in the patterns of ridership across a city following changes to cycling infrastructure. Previous studies that evaluate impacts of infrastructure improvements have primarily focused on
aspatial methods based on manual or automated cyclist counts on a single street segment. A few studies have performed retrospective analysis that aggregated data on cycling infrastructure and ridership across multiple cities and found positive correlation between cycling facilities and cycle commuting (Buehler & Pucher, 2012), but this approach does not evaluate the impact of a specific change in infrastructure. The results illustrated here demonstrate the importance of considering patterns of change in cycling when infrastructure changes in a city. In the examples of a temporary closures and installation of new infrastructure, change in one location affects the flow and amount of bicycle traffic in multiple locations. Our results suggest cyclists are shifting their routes to take advantage of the new infrastructure. As cities invest more heavily in cycling infrastructure, the need to evaluate how ridership changes is paramount.

Methodologically, the spatial weights used to implement local methods of spatial autocorrelation is an important consideration (Nelson & Boots, 2008). In network space, few studies have examined the impact of different spatial weights. Yamada and Thill (2010) mention two possible types of spatial weights matrices for networks (first order contiguity and distance based), yet they only demonstrate first order contiguity in their case study of traffic accidents in Buffalo, New York. Flahaut et al., (2003) examined the effect of higher order lags when studying traffic accidents on a highway, but their network only consisted of a single stretch of road which is not a typical network configuration. In our study, we examined the impact of several definitions of spatial weights matrices and found increased clustering with decreased neighborhood size.

Cycling data collected with GPS enabled devices represent an important source of information that is filling a massive gap in mobility data for active transportation
research. GPS cycling data has been used to gain insight into the role of cycling in meeting adults’ recommended levels of exercise and how this may be impacted by cycling infrastructure (Dill, 2009). Other studies have developed route choice models with GPS data (Broach et al., 2012). These foundational studies had small numbers of participants and gathered data for short time periods. As GPS technology has advanced and been incorporated into smartphones, mobile apps have facilitated the collection of crowdsourced cycling data. This has been picked up for research, for example, Hood et al., (2011) developed a cyclist route choice model for San Francisco, California using GPS data collected from CycleTracks, a smartphone app.

The increase in popularity of health and fitness apps, such as Strava, has provided a novel source of cycling data with high spatial and temporal density. Strava data have been used to examine where cyclists ride (Griffin & Jiao, 2015), and several studies have examined the use of Strava data as a proxy for ridership volumes (Griffin & Jiao, 2015; Jestico et al., 2016). Heesch and Langdon (2016) used heatmaps and counts of cyclists from Strava data to assess the impact of infrastructure change on cycling behavior. In the current study, we advance this field by employing Strava data to monitor spatial patterns of ridership change city-wide, and across time.

While the methods used here are well known to geographers, the application to fitness app data is an important one. A focus on detecting statistically significant change in spatial pattern of ridership is paramount to the successful use of Strava data for transportation planning and research. For example, a previous study used Strava data to evaluate the impact of infrastructure change on cycling behavior through a visual comparison of heat and volume maps pre- and post-infrastructure improvements, but
lacked a method for defining a change threshold (Heesch & Langdon, 2016). Using Local Moran’s $I$, we can determine when the change in ridership patterns are unexpected based on chance, which is a threshold that can be defended and defined statistically. Using a null hypothesis of random change gives a clear definition of change and in future studies it would be useful to evaluation null models of conditional randomness (Fortin & Jacquez, 2000).

Crowdsourced fitness App data brings new opportunities and challenges for research and practice. The unique aspect of fitness App data is that we sample movement across a city. With millions of users, Strava is an example of how fitness Apps are a growing data source and demonstrating how to effectively convert data into useful information will help fill gaps in cycling data. Like all crowdsourced data, citizen generated ridership data must be used cautiously due to inherent data biases (Feick & Roche, 2013; Ferster et al., 2017). In this and other studies, Strava data over-represent patterns of ridership in middle age males and under-represent younger and older cyclists (Griffin & Jiao, 2015; Heesch & Langdon, 2016; Jestico et al., 2016). In addition to age and gender bias, there is the potential for geographic bias, or varying uptake and use of Strava across a city (Heesch & Langdon, 2016). While bias does exist, Jestico et al. (2016) found a strong correlation between Strava and all riders in the core of a mid-sized North American city. In our study, the large user base and spatial and temporal density of the Strava data enabled a spatially explicit analysis of ridership patterns, one which has not previously been possible due to the limited spatial and temporal coverage of manual and automated cycling counts. Planning and research will continue to require official and
comprehensive count programs to monitor total number of cyclists, but the logistics of
official counts limit spatial coverage.

3.8 Conclusion

As cities continue to invest limited financial resources in cycling infrastructure,
the need to evaluate the impacts on ridership is paramount. Monitoring and evaluation of
the impacts of investment in cycling infrastructure across a city have been difficult due to
methodological issues and a lack of spatially explicit ridership data. We have
demonstrated how spatial analysis methods can be applied to emerging sources of
crowdsourced cycling data to meet this need. Our results demonstrated how change in
one location can affect flow and proportion of bicycle traffic at multiple locations. City
planners, transportation engineers and can use the methods outlined here to monitor
changes in ridership patterns following investment in cycling infrastructure or other
changes to the transportation network.
Figure 3.1 Strava users versus cycling population by age in Ottawa-Gatineau (TRANS Committee, 2011).
Figure 3.2 Definitions of spatial weights matrices for segment i: (a) first order lag, equal weighting with $w_{ij} = 1$ for contiguous street segments; (b) second order lag with $w_{ij} = 1$ for first order neighbors, $w_{ij} = 0.5$ for second order neighbors; (c) second order lag with $w_{ij} = 1$ for all neighbors.
Figure 3.3 Change in relative volume of ridership, May 2015 to May 2016.
Figure 3.4 Network local Moran’s $I_i$ of the difference in normalized ridership between May 2015 and May 2016 based on first order neighbors.
Figure 3.5 Annotated map of local network autocorrelation highlighting areas of infrastructure change during the study period.
(a) Decreased ridership on bridge shared with motor vehicles and the connecting street segments.

(b) Increased ridership on street segments connecting to the new multi-use bridge.

Cluster of increased ridership potentially due to cyclists seeking alternative routes.

Decreased ridership potentially related to tunnel closure.

Multi-use tunnel was closed at the beginning of May 2016. The tunnel was part of an important east-west cycling corridor.
Figure 3.6 Examples of annotation of changes in spatial patterns of ridership in Ottawa-Gatineau. (a) Clusters of increased ridership and decreased ridership associated with a new cyclist and pedestrian bridge. (b) Changes in patterns of ridership associated with the closure of a cyclist and pedestrian tunnel. (c) Shift of cyclists from Rideau Canal western pathway to the eastern pathway associated with temporary closures of the western pathway. (d) Shift of ridership from shared roadway bridges to multiuse path.
References


4.0 CONCLUSIONS

4.1 Discussion and conclusions

As cities invest in cycling infrastructure to increase safety and ridership levels, it is paramount that they monitor and report on the impact of new infrastructure on their citizens. Ridership levels and safety are intimately linked. Cyclists and potential cyclists frequently cite concern for personal safety as a key deterrent to cycling (Winters et al., 2011). Furthermore, as ridership levels increase, accident rates decrease (Jacobsen, 2003). Cyclists view bicycle infrastructure, such as bike lanes and cycle tracks, as safer than regular, high volume streets with no bicycle facilities (Broach et al., 2012; Hood et al., 2011) and will detour from the shortest path to their destination by about ten percent (Winters et al., 2011b). As cities continue to invest in cycling infrastructure, the need for monitoring the impacts on cycling safety and ridership levels becomes increasingly important in order to be accountable to the public and encourage political will for future investments in infrastructure (Handy et al., 2014). While monitoring the impact of new infrastructure on patterns of safety and ridership is critical, few spatially explicit approaches have been applied due to methodological limitations and a lack of data with sufficient spatial and temporal density.

In Chapter 2, we demonstrated a methodology for monitoring city-wide change in patterns of cycling safety following construction of new cycling infrastructure and applied this to a case study of the installation of cycle tracks in downtown Vancouver, Canada. First, we compared planar KDE and network KDE for mapping cycling incident intensity. Our results suggest that network KDE is more appropriate than planar KDE for
mapping cycling safety because it respects the one-dimensional nature of cycling infrastructure through the use of network distances, as well as emphasizing linear hotspots and not overestimating incident intensity. We then created maps of cycling safety in Vancouver, Canada using official reports of cycling incidents from the Insurance Corporation of British Columbia as input for network KDE. A visual comparison of the resulting maps indicated annual variation in patterns of cycling safety in addition to areas of consistent high and low incident intensity. Finally, we compared maps of cycling incident intensity before and after the installation of cycle tracks in downtown Vancouver using a change detection algorithm explicitly designed for detecting statistically significant change in kernel density estimates. We identified changes in patterns of safety defined as areas near the new cycle tracks where incident intensity increased or decreased. Additionally, we found areas unrelated to the new cycle tracks where changes in patterns of cycling safety occurred over the course of the study period. As cities continue to invest in cycling infrastructure, the need for monitoring the impacts on cycling safety becomes increasingly important, yet few spatially explicit approaches to monitoring safety exist. City planners, traffic engineers and researchers can use the KDE change detection method we demonstrated to fill this gap and monitor city-wide changes in patterns of cycling safety.

In Chapter 3 we demonstrated an approach for monitoring changes in patterns of ridership over a large geographic area using a crowdsourced cycling dataset. While standard ridership data collection occurs at a single point in time, fitness App data, such as Strava, enables mapping of ridership patterns across a city. To demonstrate how Strava data can be used to monitor ridership we created a map of the difference in normalized
ridership between May 2015 and May 2016 in the case study area of Ottawa-Gatineau, Canada. The map indicated areas of increased and decreased ridership, but it was not possible to determine if the observed change was statistically significant. We then tested for the presence of global network autocorrelation and found significant positive network autocorrelation indicating clustering of segments with increased ridership or decreased ridership. After confirming the presence of network autocorrelation, we calculated a local measure of network autocorrelation based on local Moran’s $I_i$. Our results indicate the presence of clusters of increased ridership and decreased ridership. To contextualize the results, a map of local network autocorrelation was annotated with changes that occurred to the infrastructure in the study area between May 2015 and May 2016. The annotated map indicated that changes to cycling infrastructure in one location can affect the flow and volume of cyclists in multiple locations. Overall, our results suggest that city planners, transportation engineers and researchers can use a large crowdsourced cycling dataset to monitor city-wide changes in patterns of ridership following the installation or removal of cycling infrastructure.

4.2 Research contributions

Research on the role of infrastructure on cycling safety and ridership has been gaining momentum for several decades. Early studies focused on individual segments of infrastructure, evolving to consider the role of intersections and more recently, studies are considering entire networks (Buehler & Dill, 2015). While focus has shifted towards the consideration of bicycle networks as a whole, few studies have examined city-wide patterns of safety and ridership, and even fewer studies have applied methods appropriate for network space.
Our methodology contributes to a growing body of cycling research by demonstrating the application of spatially explicit, network constrained methods for monitoring change in cycling safety and ridership. The application of network appropriate analysis methods is an important distinction. Cycling infrastructure occupies a one-dimensional subset of two-dimensional space. Traditional spatial analysis methods developed for two-dimensional space are not necessarily appropriate for network space (Miller, 1994). For example, planar kernel density estimation tends to overestimate intensity of events constrained to a network (Xie & Yan, 2008).

We have demonstrated how a traditional cycling safety dataset, comprised of incidents reported to an insurance agency, can be used to map an estimate of the intensity of cycling incidents across a city. Furthermore, we have demonstrated how the resulting maps can be compared to monitor statistically significant changes in patterns of cycling incidents following construction of new cycling infrastructure. In addition to monitoring and evaluating the impact of new infrastructure on safety, the methods can be used more generally for longitudinal monitoring of changes in patterns of safety across a transportation network. The analysis for Chapter 2 used a PostgreSQL database with the PostGIS and pgRouting extensions. The analysis was performed using custom SQL queries and Python code. All code is available for distribution upon request.

We have also demonstrated a method for monitoring changes in patterns of ridership following the installation of cycling infrastructure. Previous studies of ridership patterns have been limited by a lack of spatially explicit cycling data. Advances in GPS technology combined with an increased use of mobile health and fitness apps are changing the availability of spatially explicit data on where people ride (Jestico et al.,
2016). We have demonstrated how a large, crowdsourced cycling dataset can be used to monitor changes in patterns of ridership following changes to cycling infrastructure. Furthermore, our results demonstrate the importance of considering change in city-wide patterns of ridership as installation or removal of infrastructure at one location may affect the volume and flow of cyclists at multiple locations. The analysis for Chapter 3 used a PostgreSQL database with the PostGIS extension and Esri ArcMap 10.4. Spatial weights matrices were generated using custom SQL queries and Python code. All code is available for distribution upon request.

In addition to the manuscripts contained in this thesis, I made two significant contributions to BikeMaps.org which aided in the ongoing collection of crowdsourced cycling data. The first contribution involved the programming and publishing of a BikeMaps mobile App for Android and iOS. The second contribution consisted of ongoing maintenance and updates to the BikeMaps.org website. Collision and near miss incidents reported through the App and the website will be used in future research on cycling safety.

4.3 Research opportunities

Our methodology demonstrates spatial approaches to monitoring city-wide patterns of change in safety and ridership following changes to infrastructure, yet there are a number of limitations that require additional research.

In Chapter 2 we demonstrated a method of monitoring change in patterns of safety across a city following enhancements to cycling infrastructure. A key limitation in our analysis was the lack of exposure data. Our study calculated the absolute intensity of
cycling incidents, instead of the intensity normalized by cyclist volumes. Using our methodology it is not possible to determine if an observed hotspot represents a dangerous location, or if the incident intensity is high because a high volume of cyclists travel along that specific infrastructure. Incorporation of ridership volumes to normalize cycling incident intensity would enable the identification of changes in patterns of safety due to factors other than the number of cyclists using a piece of infrastructure. It is not unusual for studies of cycling safety to omit exposure information as detailed spatial-temporal data on ridership volumes are rarely available (Loidl et al., 2016). Crowdsourced ridership data has the potential to fill this gap in cyclist volumes. Several studies have found a strong correlation between crowdsourced cycling data and manual or automated cyclist counts (Griffin & Jiao, 2015; Jestico et al., 2016). As the research community’s understanding and acceptance of crowdsourced information increases, these data will need to be incorporated into studies of cycling safety, particularly spatial studies of safety which require data with high spatial and temporal density.

Another challenge faced by our study, and cycling safety studies in general, was a reliance on official records of cycling incidents which tend to be underreported (Tin Tin et al., 2013). Furthermore, official reports rarely record minor collisions and near-misses, both of which impact cyclists’ perception of safety (Sanders, 2015). This lack of data limits cycling safety studies. The recent development of a global tool for crowdsourced mapping of collisions and near misses, BikeMaps.org, has the potential to fill this data gap (Nelson et al., 2015). Incorporation of this emerging source of cycling safety data with our methods may aid in monitoring changes in patterns of safety following changes to infrastructure.
In Chapter 3, our analysis of the changes in patterns of ridership relies on crowdsourced cycling data which are inherently biased (Feick & Roche, 2013). There was obvious demographic bias in our ridership data, but further research is required to determine if the cycling routes travelled by the over-represented populations differ from the general population of cyclists. In addition to demographic bias, there may also be bias in the geographic distribution of the volunteer population contributing the route information (Heesch & Langdon, 2016). As such, it is unknown if the changes in patterns of ridership we observe are representative of the entire population of cyclists.

Monitoring changes in patterns of safety and ridership have been challenged by a lack of data and methodological limitations. Emerging crowdsourced cycling datasets have the potential to fill the data gap and be incorporated into spatially explicit methods for analyzing safety and ridership. Further research into the representativeness and integration of crowdsourced safety and ridership data, combined with the methods we demonstrated, will provide city planners, transportation engineers and researchers valuable tools for assessing the impact of infrastructure improvements on patterns of safety and ridership.
References


