Validation of Road Safety Surrogate Measures as a Predictor of Crash Frequency Rates on a Large-Scale Microsimulation Network

by

Alexander Ariza

A thesis submitted in conformity with the requirements for the degree of Masters of Applied Science
Civil Engineering
University of Toronto

© Copyright by Alexander Ariza 2011
Validation of Road Safety Surrogate Measures as a Predictor of Crash Frequency Rates on a Large-Scale Microsimulation Network

Alexander Ariza
Masters of Applied Science
Department of Civil Engineering
University of Toronto
2011

Abstract

A study was done to explore the suitability of intersection and arterial collision prediction models based on traffic conflicts, generated using the Paramics microsimulation suite and the Surrogate Safety Assessment Model (SSAM). A linear regression model and a generalized linear model with a negative binomial error structure were explored to correlate conflicts to crash rates, as well as the conflict-based models suggested by SSAM. The model predictions were compared to volume-based predictions and historical data from Toronto, Ontario, Canada. The volume-based predictions were calculated using a negative binomial generalized linear model, fitted to the same arterial and intersection sets used to fit the conflict-based models. The results show the predictions generated by a conflict-based model were comparable for intersections, but poor for arterials.
Acknowledgments

This thesis could not have been done without the support and advise of my two supervisors, Dr. B. Persaud and Dr. A Shalaby. Additionally, I am grateful for the crash data given to me by Alireza Hadayeghi and the simulation model maintained by Hossam Abdelgawad. I also thank my parents for all their loving support in my academic endeavours, and lastly Jess, for getting me through the writing of this thesis.
# Table of Contents

Table of Contents ............................................................................................................................... iv

List of Tables ........................................................................................................................................ vi

List of Figures ....................................................................................................................................... vii

1 Introduction ......................................................................................................................................... 1

1.1 Motivation ....................................................................................................................................... 1

1.2 Objective ....................................................................................................................................... 3

1.3 Organization of Thesis ..................................................................................................................... 4

2 Literature Review ............................................................................................................................... 6

2.1 Surrogate Safety Measures ............................................................................................................. 6

2.2 Traffic Conflict Technique ............................................................................................................. 6

2.3 Surrogate Safety Assessment Model ............................................................................................... 8

2.4 After SSAM ................................................................................................................................... 12

3 Methodology & Data ........................................................................................................................... 14

3.1 Microsimulation Calibration ........................................................................................................... 14

3.1.1 Simulated Agent to Driver Ratio ............................................................................................... 16

3.1.2 Mean Headway Time ................................................................................................................. 17

3.1.3 Mean Driver Reaction Time ..................................................................................................... 17

3.2 Conflict Aggregation ....................................................................................................................... 17

3.2.1 Positive Acceleration Conflicts ............................................................................................... 18

3.2.2 Conflict Tabulation ................................................................................................................... 19

3.3 Collision Prediction Models ........................................................................................................... 20

3.3.1 Model Forms ............................................................................................................................ 20

3.3.2 Goodness of fit ......................................................................................................................... 21
3.4 Data ................................................................................................................................. 22
  3.4.1 Crash Data .................................................................................................................. 23
  3.4.2 Microsimulation Model .............................................................................................. 23
  3.4.3 Study Area ................................................................................................................ 25

4 Microsimulation Calibration ............................................................................................... 26
  4.1 Simulated Agent to Vehicle Ratio .................................................................................. 26
  4.2 Mean Headway Time and Reaction Time ...................................................................... 29
  4.3 Shortcomings ................................................................................................................ 32

5 Model Fitting ...................................................................................................................... 34
  5.1 Arterial Models .............................................................................................................. 34
  5.2 Intersection Models ...................................................................................................... 40

6 Collision Predictions ......................................................................................................... 43
  6.1 Arterial Models .............................................................................................................. 44
  6.2 Intersection Models ...................................................................................................... 49
  6.3 Discussion ...................................................................................................................... 57

7 Conclusions ....................................................................................................................... 61

References ............................................................................................................................ 63
List of Tables

Table 1: Agent to Vehicle Calibration CPMs ................................................................. 27
Table 2: Driver Behaviour Calibration CPMs ................................................................. 30
Table 3: Fitted Arterial Conflict-Based CPMs ............................................................... 36
Table 4: Fitted Arterial Volume-Based CPMs ................................................................. 37
Table 5: Revised Arterial CPM ...................................................................................... 40
Table 6: Fitted Intersection Conflict-Based CPMs ......................................................... 40
Table 7: Fitted Intersection Volume-Based CPMs ......................................................... 41
Table 8: Arterial Crash Prediction Goodness of Fit ....................................................... 44
Table 9: Intersection Crash Prediction Goodness of Fit ............................................... 50
List of Figures

Figure 1: Map of the Study Area ................................................................. 24
Figure 2: Screenshot of Microsimulation Model ........................................... 24
Figure 3: Arterial CURE Plot for Volume-Based CPM ................................... 45
Figure 4: Arterial CURE Plot for SSAM Linearized CPM ............................... 46
Figure 5: Arterial CURE Plot for SSAM Non-Linearized CPM ....................... 46
Figure 6: Arterial CURE Plot for Linear Conflict-Based CPM ....................... 47
Figure 7: Arterial CURE Plot for NB Conflict-Based CPM ............................ 47
Figure 8: Arterial CURE Plot for NB Length and Conflict-Based CPM ............ 48
Figure 9: Major AADT Intersection CURE Plot for Simulation-Fitted Volume-Based CPM 51
Figure 10: Minor AADT Intersection CURE Plot for Simulation-Fitted Volume-Based CPM 51
Figure 11: Major AADT Intersection CURE Plot for City-Fitted Volume-Based CPM 52
Figure 12: Minor AADT Intersection CURE Plot for City-Fitted Volume-Based CPM 53
Figure 13: Intersection CURE Plot for SSAM Linearized CPM ..................... 54
Figure 14: Intersection CURE Plot for SSAM Non-Linearized CPM ................ 54
Figure 15: Intersection CURE Plot for Linear Conflict-Based CPM ................ 55
Figure 16: Intersection CURE Plot for NB Conflict-Based CPM .................... 56
Figure 17: Intersection CURE Plot for Linear and NB Conflict-Based CPMs .... 56
1 Introduction

The movement of goods and people is essential to modern society. Freight is produced in locations far removed from where they are consumed, and people commute every day in order to get to work. As the global population continues to grow, the population of urban areas also increases. More people leads to more traffic and more accidents on the road. The WHO reports that 1.3 million people die annually due to road traffic accidents. Current projections show that road crashes will become the fifth leading cause of death by 2030. As a result, traffic engineers continue to seek out models to explain the effects of various factors on the safety performance of transportation networks.

1.1 Motivation

The typical method of evaluating the safety of a roadway is based on historical crash records. Currently, a generalized linear crash prediction model is used to correlate factors such as geometric features and traffic volume to crash rates at a given location. These models take these factors as explanatory variables and produce an expected crash frequency. Unfortunately, there are many problems with relying on crash records. For one, accidents occur infrequently and sporadically, so many years of historical data are required for the crash frequency data to be useful. The historical data that does exist for crashes is usually incomplete, and minor crashes are often unrecorded (Parker and Zegeer, 1989). As a result, the predictions generated by volume-based models may be based on incomplete data. Additionally, historical models are only possible where crash records exist; newer intersections and road segments can only be evaluated for safety indirectly.

The factors used in the formulation of a volume-based model are limited by the need to code them into numeric values. For example, a factor such as roadway curvature must be coded and calculated for each roadway segment to be analyzed. If most of the road segments that are used to fit the model are straight segments, it is possible that the model fitting will find that this factor is not statistically significant. It is also possible that too many explanatory variables can be included when forming a model, resulting in the
model becoming over-fitted, and providing less valuable predictions for another set of intersections or arterials.

Due to the fact that these volume-based models are calibrated to historical data, it is impossible to test the impact of some types of proposed projects. In the case of a proposed roadway, it may be possible to generate predicted collision rates based on the roadway characteristics and the projected traffic volumes. However, if a proposed project will not affect any of the model variables, the model predictions will be unchanged. For example, if a roadway were to implement variable speed limits, a volume-based model would fail to project a change in crash frequency unless the model included variable speed limits as a factor. If this roadway were the first of its kind in an urban area, it would be impossible to calibrate a model to the driving behaviour of the local drivers, and thus, the safety impacts of variable speed limits could not be evaluated prior to implementation.

In order to overcome these problems surrogate safety measures such as traffic conflicts are used. Traffic conflicts can be observed in real life, but they can also be estimated in microscopic simulations of road traffic. A project was undertaken by the Federal Highway Administration to produce a framework for analyzing the vehicle paths during a microsimulation run. Years of development have led to a program to be used in conjunction with simulation software suites, which post-processes the simulation’s data output to identify and list all conflicts. These conflicts have been related to crashes at intersections, but it is not known if they can predict crashes for arterials, or if the established relationships are the most robust ones.

Microsimulation models are becoming more and more commonly used for the planning and design of transportation projects. Conflicts can be extracted from models that are being used for the development of infrastructure projects, and if conflicts can be used to predict crash frequencies, projects that are planned and designed with microsimulation models can easily be evaluated on their safety impacts. Conflict-based predictions could allow for the ability to identify the safety impacts of proposed projects. Considering the
cost of roadway projects, both financially and in time, it would be ideal to have a more complete understanding of both the impact of a project and its possible alternatives.

1.2 Objective

This study aims to both test the predictive ability of conflict-based collision-prediction models for arterials and intersections, and test their suitability for arterial segments. The validation process employed by SSAM and subsequent papers focused exclusively on intersection cases, not arterial road segments. During the SSAM validation it was found that surrogate measures generated by microsimulation models are significantly correlated with historical crash data. The traditional crash prediction models (CPMs) based on average daily traffic did perform better and have higher correlation to historical data than the traffic conflicts generated by SSAM. Papers published since the release of SSAM have included its application to intersections (Zhou et al, 2009; Li and Sun, 2010). However, no effort has been with the goal of validating the SSAM approach to highway traffic and arterial incidents.

The microsimulation-based safety analyses for non-intersection cases have looked only at the conflicts generated, or other surrogate measures. No comparisons were made between the conflicts on arterials and the historical crash records, or crash predictions. Liu and Garber (2007) used the framework presented by the FHWA in 2003 to fit the case of a highway. They were able to apply conflicts to a highway case, but the conflicts were not compared with real-world crash data. Some research is being performed with the goals of deriving simulation-based safety indicators for highways, such as the work by Cunto et al. (2009); however, their method uses a crash potential index and does not predict crash frequencies. No effort has been made to compare conflicts from microsimulation to historical crash data for cases outside of intersections.

As it stands, the correlation between simulated conflicts and crashes is untested for arterial segments. It is possible to predict crash frequency using volume-based models for
both arterials and intersections, but it is not known how those predictions compare to predictions generated by simulated conflict-based models.

The objectives of this thesis are:

- Testing the correlation between conflicts and crashes for arterial segments
- Investigating the transferability of the SSAM conflict-based crash prediction model to other study areas
- Evaluating the predictive capabilities of conflict-based models in comparison with traditional volume-based models, for both intersections and arterials

1.3 Organization of Thesis

The study undertaken consists of three parts. The first part describes the calibration of a microsimulation network with the goal of producing the best conflict-based crash predictions. In this phase, various simulation factors are investigated for their effect on crash predictions. After each simulation run, crash rates are predicted based on the conflict counts during each simulation, and compared to the historical crash rates. The set of parameters that produce crash predictions that best fit the historical crash counts are used for the third phase.

The second part documents the development of crash-prediction models based both on conflicts and flows, in order to determine the viability of conflict-based crash predictions for use in the real world. The testing required the calibration of two types of crash prediction models, one based on conflicts and one based on vehicular flows. These prediction models were developed for intersections and arterials. The conflicts and vehicular flows come from the same simulation model, so that the two prediction models can be compared to historical data.

The third part provides a comparison between the predictions from conflict- and volume-based models. Predictions are generated based on the flows and conflict counts for a set of intersections and arterial segments from the simulation model, but not the same set used to calibrate the models. Both conflict and volume-based models are calibrated using the set of intersections and arterials that were used in the second phase. Since the
predictions are based on simulated traffic volumes and simulated conflicts, and the models are all calibrated using the same set of intersections and arterials, it is possible to make a direct comparison as to the usefulness of volume vs. conflicts as an explanatory variable in prediction collision frequencies.

This thesis is broken down into 7 chapters. Chapter 2 presents a review of the literature leading up to the development of SSAM, and what has come since. Chapter 3 specifies the methodology that is followed, as well as the data sources used in this study. Chapter 4 presents the results and observations from the microsimulation model calibration phase. Chapter 5 describes the crash-prediction model calibration. Chapter 6 provides the results of the crash prediction models and their implications. Finally, chapter 7 concludes this work.
2 Literature Review

2.1 Surrogate Safety Measures

Surrogate safety measures are any events that can be correlated with crash rates. Because these methods use events that occur at a much greater frequency than crash rates, it is possible to assess the safety of a given location without waiting for a large number of crashes to occur. Additionally, these measures can be used with microsimulated road networks to assess the safety of proposed roadways and transit projects, experimental roadway designs, or operational strategies before they are built or implemented.

A number of factors have been proposed for use as surrogate safety measures, such as volume, speed, delay, accepted gaps, headways, shockwaves, and deceleration-to-safety time (FHWA, 1981). Other potential surrogates are critical events, such as lane merging, speeding, and running red lights (Kloeden et al., 1997; Porter et al., 1999). The most frequently used surrogate measure is traffic conflicts. A traffic conflict is the occurrence of two (or more) road users that risk colliding if their course is unchanged.

2.2 Traffic Conflict Technique

Traffic conflicts can trace their history back to 1954, when McFarland and Moseley first reported their observations of “near misses”, or “emergency situations or critical incidents which could easily have led to an accident”, experienced by buses and trucks in an urban setting. The first systematic attempt to record traffic conflicts was in 1966 by two General Motors engineers. The pair found that most drivers reacted to potential conflicts by engaging in an aggressive maneuver, such as braking or lane-changing (Perkins and Harris, 1969). The traffic conflict technique was shortly thereafter adapted to evaluate crash potential at intersections. When traffic conflicts were first being recorded, they were observed and recorded by field investigators, looking for strong breaking and evasive manoeuvres at intersections (Parker and Zegeer, 1989).

A number of authors have aimed at validating the traffic conflict technique. Glauz et al. (1985) aimed at showing a correlation between conflicts and crashes. They collected conflict and crash data at 46 urban intersections, producing tables of daily conflict rates
for intersection based on traffic volume and signal control. The validation of traffic conflicts was done for each class of intersections. Within each class, two intersections were randomly selected, and the expected accident rates and their variances were calculated using the tables of conflict rates. The expected accident rates based on conflicts was 18.20, and the total number of accidents that was predicted with traditional means was 20. As a result, the authors concluded that the crash predictions from conflicts are equally valid compared to those predicted from historical data.

Field observers were a source of error when collecting conflict data, due to the subjective nature of deciding if a given driving event is a conflict or not. Each observer is required to judge whether or not a situation is a conflict, resulting in variability in the grading of traffic conflicts by different people. As a result, the human-collected data was not necessarily accurate, especially if multiple observers were used. Nonetheless, traffic conflicts have been shown to have some correlation with crash frequency, and the consensus is that higher rates of conflicts correlate to lower levels of safety (Gettman et al., 2008).

To eliminate the subjectivity from traffic conflict analysis, objective measures are used. The most common of these measures is time-to-collision (TTC), which is defined as the time until two vehicles on a collision course do collide, if they continue on their current trajectories (Hayward, 1972). As vehicles continue on a collision course the TTC decreases, so it is the minimum TTC that is critical. Based on this, conflicts can be classified and sorted by severity. Lower minimum TTC values leave less time to carry out corrective action, the more severe the collision. Many other objective measures exist, such as post-encroachment time. Post-encroachment time (PET) is the time between two vehicles on a near-collision course passing at a common point (Allen et al., 1978, van der Horst & Kraay, 1986). As with TTC, a lower PET indicates higher severity, and the minimum value is also the critical value.

Sayed and Zein (1999) developed regression prediction models correlating traffic conflicts defined by TTC to traffic volumes and crash rates. Both conflicts and crashes were assumed to follow a Poisson relationship. A statistically significant correlation was
found between crashes and conflicts, with an R-square value ranging between 0.70 and 0.77 for signalized intersections. No significant relationship was found at unsignalized intersections.

Once objective measures for traffic conflicts were devised, it became possible to measure traffic conflicts in computer software. Archer and Kosonen (2000) were among the first to investigate the possibility of using microsimulation to perform safety analysis. The Federal Highway Administration (FHWA) released a report in 2003 entitled: “Surrogate Safety Measures From Traffic Simulation Models”. This report marked the start of a concerted effort to extract surrogate safety measures from microsimulation models.

The FHWA approach records all conflict events between any two vehicles during the entire simulation run. There are two criteria for a conflict: a vehicle in simulation must engage in an evasive action, and the resulting surrogate measure must be under a predefined threshold. A number of surrogate measures were proposed: time to collision (TTC), post-encroachment time (PET), maximum speed of the two vehicles, maximum difference in speed between the two vehicles during the conflict event, initial deceleration rate (DR) of the reacting vehicle, and the location of the start and end points of the conflict event. The report included algorithms to calculate the surrogate measures for different conflict types. In order for surrogates to be extracted and computed from microsimulation, the major microsimulation suites need to be able to output data that can be input into a post-processor. The format proposed for this data was an “event file” that would contain a time history of the speed, acceleration, and location of vehicles that are participants in a conflict during the simulation run. The event file would be created as a plug-in to simulation software and then separately post-processed. This process was named the Surrogate Safety Assessment Model (SSAM).

2.3 Surrogate Safety Assessment Model

In 2008, following in the footsteps of their 2003 report, the FHWA developed a post-processing software utility to analyze traffic conflicts from traffic simulation. The SSAM utility is compatible with four traffic microsimulation suites: AIMSUN, PARAMICS, TEXAS, and VISSIM. The microsimulation suites are able to generate a trajectory (TRJ)
file for each simulation run. The TRJ file is used as an input into SSAM, which then identifies traffic conflicts that have occurred during the simulation run. The conflicts are classified by type of conflict, and surrogate safety measures are computed.

Validation of SSAM included three steps: a sensitivity analysis, theoretical validation, and field validation. The sensitivity analysis was performed to ensure that the results of running SSAM on the same model would not vary significantly based on which microsimulation suite generated the TRJ file. The four major microsimulation suites led to different results, but not differences of any significance. The theoretical validation was done with a series of different intersection layouts in simulation models. This was done to ensure that SSAM could differentiate between different intersection designs, and secondarily to identify correlations between existing intersection models and the surrogate safety measures produced by SSAM. The field validation was done to correlate the conflicts produced by microsimulated real-world sites with historical crash data for those sites, as well as comparing the conflict rates to traditional crash prediction models. The field validation consisted of five tests designed to evaluate how well the microsimulation models would compare to analysis done with historical crash data.

The first test was to compare a ranked set of intersections that were ordered by two criteria: annual crash frequency from historical data, and average hourly conflict frequency from microsimulated data. A set of intersections was modeled and conflicts computed using SSAM, with each model running for a simulation time of one hour, and each simulation run five times. The average numbers of conflicts per hour were computed by averaging each of the five simulation runs, providing the average hourly conflict frequency to order the set of intersections. The set of intersections all had a minimum of three years of crash data used to order the set by crash frequency. The two intersection ranked lists were compared and the Spearman rank correlation coefficient was computed to be 0.463, which is significant at the 95-percent level of confidence. Therefore, a significant correlation exists between the conflict frequency and crash frequency. For the purposes of this test, the simulation was run using AM-peak values, and not average daytime traffic values, so it is possible that a closer correlation could have been found were the simulations more comprehensive.
The second test was a comparison of a ranked set of intersections that were grouped by three incident types, specifically rear end, crossing, and lane changing. Upon computing the conflicts per hour and comparing them to the number of crashes per year of each type of incident, varying distributions were found between each of the types of incidents. When comparing conflicts per hour to crashes per year, the ratio was found to be 0.01 for rear end incidents, 2.06 for crossing incidents, and 0.65 for lane changing incidents. The Spearman rank correlation coefficient value for the rear end incidents was 0.473, and for lane changing incidents it was 0.469; both significant at the 95-percent level of confidence. The number of crossing conflicts was too low for a Spearman rank correlation coefficient to be calculated. The goal of this test was to determine if SSAM could accurately identify intersections prone to a specific crash type. The test did find a significant correlation between rear end conflicts and crashes, as well as lane changing incidents and crashes. The conflicts-to-crashes ratio varied based on the type of incident, which the FHWA attributes to the possibility that the conflicts-to-crashes ratio is lower for more severe incidents.

The third test was to correlate conflicts to crashes. This was done by defining an equation to estimate average yearly crash frequencies at an intersection as a function of hourly conflict frequencies. A regression equation was used to develop this conflict model for crash prediction. The regression model was developed and three statistical goodness of fit measures (Pearson chi-squared, scaled deviance, and R-squared) were used to calibrate the model to historical crash data. Two models were developed, one linear (Equation 1) and one non-linear (Equation 2). The models related Crashes, in terms of average yearly crash frequency with Conflicts, expressed in hourly conflict frequency. The linear model was a standard linear regression model, and it had an R-square value of 0.27. The non-linear model is a generalized linear regression model with a negative binomial error structure, and had an R-square value of 0.41, using Miaou’s pseudo R-square value (Miaou, 1996). Although both models had low R-squared values, these values are still within the range of traditional crash prediction models.

\[
\log(\text{Crashes}) = 1.09 \times \log(\text{Conflicts}) - 0.98 \quad \text{Equation 1}
\]
The models were compared against a traditional volume-based crash-prediction model, prepared using a generalized linear modeling approach. In this case, the average yearly crash frequency was modeled as a function of the average annual daily traffic (AADT). The SSAM study found an R-square value of 0.68 for the volume-based model. In their study, the volume-based model had a better correlation to crash data than the conflict based model. The volume-based model was based on volumes from historical data, whereas the conflict counts were based on simulated volumes, so the two models were fitted on different data sets. The two models were compared only on the basis of their fit, and not for their predictive abilities. Additionally, the validation effort focused only on intersections, and not arterial roads.

The fourth test was to identify locations that are prone to crashes. A conflict prediction model was developed to predict hourly conflict rates as a function of traffic volume. A traditional crash prediction model was also developed to predict yearly crash frequency as a function of traffic volume. The two models were then used to identify incident prone locations (those with more incidents than a predefined “normal” level based on average incident rates). The crash prediction model identified 20 crash-prone intersections, while the conflict prediction model identified 12 conflict-prone locations, and only one location was identified by both. Therefore, there is no agreement between the conflict and crash models in identifying incident prone locations. When the lists of sites were ranked in terms of priority for improvement using a predicted/normal incident ratio and a “potential for improvement” scheme, the Spearman rank correlation coefficient values were 0.001 and 0.033, implying an insignificant correlation between intersections with excessively high numbers of crashes and those with an excessively high number of conflicts. The high degree of difference between the two models was explained by the fact that the two models induce opposite biases in the ranking of sites for improvement. The concavity of the crash-prediction model is upward, while the conflict-prediction model is concave down. The report finds that this test may not be conclusive due to the differing nature of the conflict and crash prediction models.
The fifth test was the same as the fourth, except performed for each conflict type (rear-end, lane-changing, and crossing). No significant correlations were found between the conflict and crash based predictions for any type of incident. The reasoning for this is the same as with the fourth test.

2.4 After SSAM

The validation of SSAM found that the surrogate measures generated by microsimulation models were significantly correlated with historical crash data. The traditional crash prediction models based on average daily traffic did have a higher correlation to historical data than the traffic conflicts generated by SSAM. Nonetheless, the correlation between simulated conflicts and historical crashes is significant. Additionally, the validation effort focused only on intersections, and not arterial roads, highways, or networks of roads.

Further studies have been performed using SSAM, exclusively focusing on intersection scenarios. Zhou et al. (2009) analyzed the conflicts that resulted from a simulated intersection, modeled both in its original state and after improvements are made to the intersection. In their study, they found that the simulation after improvement showed decreased numbers of crossing, rear-end, and lane-change conflicts. They found microsimulated conflicts useful in evaluating the effect of road improvements on safety.

Li and Sun (2010) further demonstrated the use of SSAM for intersections, this time proposing a two-stage method to calibrate microsimulation parameters. A set of three intersections are used and modeled. Calibration results show that parameter values have a significant effect on the accuracy of conflicts generated by the simulation. Additionally, common calibration methods need to be carefully scrutinized since they may result in behavioural parameter values that do not reflect actual driver behaviour.

Dijkstra et al. (2010) determined conflicts in a large-scale microsimulation model, looking specifically at intersections. Their statistical model correlated both number of conflicts and flow with collisions, and concluded that there is a correlation between simulated conflicts and collisions. The authors did not use SSAM but rather developed
their own set of algorithms to identify conflicts. Their conflicts were governed by a max TTC, just like SSAM.

SSAM was initially designed with intersections in mind, so intersection safety was clearly the focus of the SSAM validation effort. However, no papers have been published with the goal of validating the SSAM approach to highway traffic and arterial incidents. Furthermore, no research has been done to further validate SSAM, or test its fitness for areas other than intersections. SSAM works by identifying all conflicts that occur during a simulation run, and therefore also returns conflicts that occur outside of intersections. Little validation effort has been expended to verify if these conflicts also correlate to crashes.

Liu and Garber (2007) used the framework presented by the FHWA in 2003 to fit the case of a highway for the purposes of examining the impact of truck-lane restriction strategies. Their study evaluated the conflicts from a highway segment, with various restriction schemes in effect. The study illustrated the safety implications of different operational strategies on a highway segment; however, the microsimulation results were not compared with a real-world application.

Cunto and Duong (2009) endeavour to link crash counts with simulations for highway segments. Their work utilizes a crash potential index, based on vehicle deceleration rates. Their work shows that crashes tend to occur when the crash potential is high. The crash potential index does not provide any method to predict crash frequencies, but only indicates that there is some evidence that measures of safety performance can correlate to crash frequency.

In summary, SSAM allows the identification of conflicts, and conflicts have been shown to correlate to crashes when analyzing intersections. Conflicts can be observed occurring in simulations at arterial and highway segments, and microsimulation has been used for the purposes of safety analysis in these cases. However, these conflicts have not been tested to see if there exists a correlation between these conflicts and crashes.
3 Methodology & Data

The first part of the study was the calibration of a microsimulation model. Once calibrated parameter values were chosen, the next step was to run the model a number of times to provide a set of average conflicts and flows. The output from these simulation runs were used to generate both volume- and conflict-based crash-prediction models, and the models were compared. The volume- and conflict-based models were fit using only half of the simulation arterials and intersections, chosen at random, so that the remaining half could be used to generate predictions using the fitted models. Once predictions were generated using both models, both sets of predictions were compared to the historical crash counts for each location.

3.1 Microsimulation Calibration

The microsimulation modelling software used was Paramics. Paramics is an agent-based microscopic traffic simulation suite. Vehicles are modelled as individual agents whose movements are governed by car-following, lane changing, and gap acceptance models. In order to allow the user the opportunity to adjust the behaviour of the drivers, there are a number of adjustable parameters. These parameters, in conjunction with a road network, are the inputs to Paramics, which produces vehicle movements as an output. Parameter values must be adjusted in order to produce vehicle movements that match those observed in the area being modelled, otherwise the simulation will not produce results consistent with the real world.

Calibration of any microsimulation requires some objective. Often, microsimulation models are calibrated to most closely match the flows of the network being modeled. The objective in such a case would be to produce a set of traffic volumes that most closely match the traffic counts in real life. Additional outputs, such as queue lengths and travel times, are also often checked. For the purposes of this study, the simulation model has already been calibrated for a previous work (Gardiner-Lake Shore Scoping Study, 2003), and is now being applied to generate predictions of crash frequencies. Consequently, the calibration objective of this study is to generate predictions that most closely match the historical crash data for the model area.
In order to generate crash predictions, two model forms suggested by SSAM were used. A linearized conflict-based crash-prediction model was used to generate crash predictions, which was compared with historical data. The linearized form is shown in Equation 3. The values of $\alpha$ & $\beta$ were estimated by using those values that minimize the residual sum of squares.

$$\log(\text{Crashes}) = \alpha \times \log(\text{Conflicts}) - \beta$$

Equation 3

To evaluate how well the predicted crash counts match the historical crash data, two goodness of fit measures were computed: the R-squared value, calculated as shown in Equation 4, and the Pearson chi-squared value, shown in Equation 5. These two statistics are easily calculated, and provide a simple way to compare the predictions generated from different simulation parameters. Since the goodness of fit measurements were used only to indicate if a set of parameters represented an improvement over the previous set of parameters, and merely guided the parameter choice, more complex goodness of fit measurements were not required.

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}}$$

Equation 4

where $SS_{err} = \sum_i (O_i - E_i)^2$, the residual sum of squares

$SS_{tot} = \sum_i (O_i - \overline{O})^2$, the total sum of squares

$O_i$ are the observed crash counts

$\overline{O}$ is the mean crash count

$E_i$ are the predicted crash counts

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Equation 5

where $O_i$ are the observed crash counts

$E_i$ are the predicted crash counts

The calibration approach was a trial-and-error method. More complex and comprehensive calibration methods exist, using computerized optimization techniques to
search for an optimal set of parameters. Such methods require a fully automated analysis process. Unfortunately, SSAM only provides a graphical user interface, so all conflict analyses must be entered by hand. The inability to fully automate the process, coupled with the long analysis computation time, led to the decision to calibrate simply by trial-and-error.

A comprehensive testing of many parameters would have become overly tedious, so only a few parameters were chosen for calibration. The parameters that were adjusted are simulated agent-to-driver ratio, mean headway time, and mean driver reaction time. In order for Paramics to implement the mean driver reaction time accurately, the speed memory and time steps factors needed to be adjusted accordingly. The first factor affects the resolution of the simulation model. The other two parameters, mean headway time and reaction time, are factors which influence driving behaviour.

3.1.1 Simulated Agent to Driver Ratio

The number of vehicles that one simulated agent represents alters the results of the simulation, as more simulated vehicles results in more accurate modeling of traffic. The network simulated was originally set with a vehicle divisor of 100, that is, one simulated agent is released onto the network for each 100 vehicles in the origin-destination matrix. The origin-destination matrix was set at real-world levels, so this effectively is the simulated agent to driver ratio.

Ideally, one would carry out a simulation at a one agent to one driver ratio. However, the computational resources required to simulate a model, especially a large network, are overly high, and the resulting output file sizes also increases with the number of simulated vehicles. The initial simulation run occurred with a ratio of one agent per hundred vehicles, and the divisor was subsequently decreased until either there was no noticeable gain in predictive ability, or the computational demands became too great to process with a single computer.
3.1.2 Mean Headway Time

The mean headway time (MHT) is the average time between successive vehicles in the simulation. As simulated drivers drive closer together, the mean headway time decreases. One of Paramics adjustable factors is the mean target headway. When this coefficient is increased, Paramics will adjust the simulated behavior of drivers in order to space the vehicles farther apart, conversely, reducing the mean target headway causes simulated drivers to drive closer together, increasing the capacity of the roadway. Headway is related to gap, which is the distance between successive vehicles. Paramics operates as a gap acceptance model, so vehicular agents will behave based on the minimum headway that it is programmed to accept. For example, if an agent finds that the headway is approaching the minimum headway permitted, it will either slow down or change lanes. The lower the headway, the more aggressive the agents will drive. Aggressive driving can result in some road segments and intersections becoming more dangerous than others, therefore the mean headway value needs to be adjusted in order to most accurately simulate the driving behavior in the study area in order to obtain conflicts that can be correlated to crash rates. The value of mean target headway suggested by Paramics is 0.85 to 0.90 seconds for urban areas.

3.1.3 Mean Driver Reaction Time

The mean reaction time (MRT) is defined in Paramics as the average time between the observation of a change in the speed of the vehicle ahead of a simulated driver, and that driver’s reaction to the change in the preceding vehicle’s speed. As the reaction time decreases, driving behavior becomes more aggressive. In order to accurately implement the target mean driver reaction time, the value of the speed memory is recommended to be set to 150% of the mean reaction time, in terms of time steps.

3.2 Conflict Aggregation

Once the microsimulation model parameters are chosen, the simulation is performed. The output from Paramics is a set of TRJ files, representing the position of all vehicles during each time step of the simulation, as well as traffic volumes on each link. The TRJ files are input into SSAM and processed. The output from SSAM is a list of all conflicts.
which occurred during the time represented by the input TRJ files. For a simulation of one hour, the number of conflicts that occurs in the model area is on the order of hundreds of thousands, and each conflict is one line item. In order to aggregate the individual conflicts into conflict counts for intersections and arterial segments, some work needed to be done and some assumptions needed to be made. It was observed that some conflicts had a positive acceleration time, and they could either be removed as false conflicts, or rationalized to explain their inclusion. Another set of false conflicts arises from roadways that cross at different elevations, such as overpasses. SSAM does not consider elevation in its analysis, so vehicles at different elevations that cross will be identified as conflicts. These false positives were removed as conflict counts are tabulated for each intersection and arterial segment.

3.2.1 Positive Acceleration Conflicts

Some runs were re-analyzed, but any conflicts with a positive acceleration time were removed. It has been suggested that conflicts where the following vehicle has a positive acceleration rate is a false positive. This case of a false positive conflict follows from the idea that if a vehicle is in danger of colliding with another vehicle, the following vehicle will slow down and attempt to avoid collision. The Paramics car-following model uses this assumption, and the mean target headway is the factor that dictates how closely a following vehicle will get before slowing down. An agent will only accelerate if the current gap between itself and the next car is great enough that accelerating will not cause the following agent to decrease to an unacceptable headway. As a result, in all arterial conflicts, the following vehicle will be decelerating, and all accelerating conflicts are false positives for arterial analysis.

There is a case of conflicts where it is possible that the two vehicles will be accelerating towards each other, namely, crossing conflicts. The car following model of Paramics makes no provision for vehicles crossing paths with each other too closely, especially at intersections. Vehicles at intersections do accelerate as they enter the intersection, and it is quite possible that a crossing conflict will occur at an intersection where the following vehicle accelerates, not sensing the possibility of engaging into a conflict situation.
Therefore, for the case of intersection analysis, accelerating conflicts are not considered false positives and are included in the results.

The inclusion of conflicts where the following vehicle has a positive acceleration was supported by the better fit of a set of conflicts including the accelerating conflicts as compared to the same set of conflicts with the positive acceleration conflicts removed.

### 3.2.2 Conflict Tabulation

Once objectionable conflicts were removed, the remaining conflicts were tabulated by the link of the first vehicle, and the link of the second vehicle. A count of the number of total conflicts, lane change conflicts, rear end conflicts, and crossing conflicts was produced for each first link-second link pair.

One flaw in the way that SSAM operates is that it does not recognize different roadway elevations in the network. As a result, a vehicle on an elevated roadway and a vehicle on a roadway underneath may be identified as conflicting numerous times, while in reality, no conflict occurs (Gettman et al., 2008). Since the conflicts were counted by first link-second link pairs, it was trivial to check to see if the first and second links intersect in the model. If not, it was a result of links that cross at different elevations.

The link lookup provided one other function, identifying if those conflicts occur on an arterial segment or at an intersection. A list of intersections was prepared where each intersection included a list of the links flowing into that intersection. Similarly, a list of arterials was prepared with each entry having a list of link segments with which the arterial segment was comprised. The two initial links for each conflict were compared with these lists and attributed to the intersection or arterial segment that the conflict occurs on. During this processing, the traffic volumes from Paramics were also attributed to the intersection or arterial segment that link represents, in order to provide hourly volumes.
3.3 Collision Prediction Models

3.3.1 Model Forms

The two model forms to be used for generating crash predictions from conflicts were the same ones used to generate the predictions for the calibration phase: the linearized model used in the calibration phase (Equation 3), and a generalized linear model with a negative binomial structure (Equation 6). The linear model was fitted using an ordinary least squares method, while the other model was fitted using a generalized linear model (GLM) fitting procedure with a negative binomial error distribution, computed using the R statistical computing environment. Models were created both for arterials, and intersections.

\[ \text{Crashes} = \alpha \times \text{Conflicts}^\beta \]  \hspace{1cm} \text{Equation 6}

The volume-based collision prediction models were negative binomial linearized models. Two models were used, one for arterials (Equation 7) and one for intersections (Equation 8). As with the negative binomial conflict-based model, the volume-based models were fitted using a GLM procedure in the R computing environment.

\[ \text{Crashes} = L^{a_1 \times AADT^{a_2} \times \exp (\beta)} \]  \hspace{1cm} \text{Equation 7}

where \( L \) is the length of the arterial segment

\( AADT \) is the average annual daily traffic

\[ \text{Crashes} = AADT_{maj}^{a_1} \times AADT_{min}^{a_2} \times \exp (\beta) \]  \hspace{1cm} \text{Equation 8}

where \( L \) is the length of the arterial segment

\( AADT_{maj} \) is the average annual daily traffic on the major arterial

\( AADT_{min} \) is the average annual daily traffic on the minor arterial

The volume-based models used traffic volumes in the form of AADT, but the traffic count output from the microsimulation was hourly. In order to convert the hourly figure
HV into an average daily traffic flow, the conversion shown in Equation 9 was used. Equation 9 required a conversion factor \( K \), which was found in the Highway Capacity Manual (2000). For urbanized areas, \( K = 0.091 \).

\[
AADT = \frac{HV}{K} \quad \text{Equation 9}
\]

### 3.3.2 Goodness of fit

In order to evaluate the calibrated collision prediction models, goodness of fit measures were required. For the negative binomial models, three goodness of fit measures were calculated: Pearson chi-square, scaled deviance, and Miaou’s \( R^2_\alpha \). These three parameters are jointly used in the evaluation of goodness of fit of the volume- and collision-based models.

The Pearson chi-square, calculated as shown in Equation 5, can also be converted into the mean Pearson chi-squared by dividing it by its degrees of freedom. The mean Pearson chi-squared should have a value between 0.8 and 1.2 for the negative binomial error distribution to be appropriate. Deviance is defined in Equation 10, and the scaled deviance is the deviance divided by the model’s scale parameter. The mean scaled deviance is the scaled deviance divided by the degrees of freedom. The mean scaled deviance should be approximately 1 if the model provides a reasonable fit of the data. A mean scaled deviance larger than one indicates over-dispersion in the data, while a smaller value indicates under-dispersion.

\[
D(y) = -2 \left[ \log L(\hat{\theta}_0) - \log L(\hat{\theta}_s) \right] \quad \text{Equation 10}
\]

where \( \hat{\theta}_0 \) is set of fitted values

\( \hat{\theta}_s \) is set of values from a saturated model, one where a parameter is estimated for each observation

Miaou’s \( R^2_\alpha \) is used to evaluate how well the variance of the data is explained by the model, in comparison to a model with no explanatory variables, and only a constant term. The calculation is shown in Equation 11. The dispersion parameter, \( \alpha \), indicates
\[ R^2_\alpha = 1 - \frac{\alpha}{\alpha_{max}} \]  

Equation 11

where \( \alpha \) is the dispersion parameter for the model
\( \alpha_{max} \) is the dispersion parameter for the model with
only a constant term

The linear conflict-based prediction models were evaluated using the R-square and Pearson chi-squared values. A higher R-square value indicates a better fit. For the Pearson chi-squared value, a lower value indicates a better fit.

Once the crash prediction models were calibrated, predictions were generated. Conflicts and volumes from a set of simulated arterials and intersections were fed into the models. To evaluate the predictions presented by each model, the mean Pearson chi-squared value and the R-square statistic were calculated. The quality of the predictions from each of the models was also evaluated using cumulative residual (CURE) method (Hauer & Bamfo, 1997). CURE plots were created for the volume-based models by plotting cumulative residuals against AADT, and against conflicts for the conflict-based models. The CURE plots allowed a graphical representation of the fit of each model.

CURE plots display the cumulative residuals, as well as two standard deviations across the entire AADT or conflict range. In a good model, the cumulative residuals will oscillate about 0, and remain within the bounds of two standard deviations. A poor model will frequently exceed two standard deviations, or will remain constantly to one side of 0.

3.4 Data

The data required for this thesis were crash counts for a five-year period, and a microsimulation network. Ideally, the microsimulation network would exactly reflect the time span represented in the historical network, however, the data used in this simulation slightly predates the microsimulation network. The crash data was collected over a five year period, while the microsimulation represents the road network a year later. Slight changes to the road network may have occurred between the period the crash data was collected and the state of the network as modelled in the simulation, however, no major
changes occurred during this time. Major changes would include items such as new road
construction, permanent road closures, and reconfigured intersections. It is probably that
more minor details, such as traffic signal timing, did change between the crash data
collection and the microsimulation network, but these changes would also be present
during the five year period in which crash data was recorded. As a result, these minor
changes are not significant.

3.4.1 Crash Data

Historical data for the model area was provided from the City of Toronto. Crash data for
major arterials in the study area was split into crash counts for 2411 arterial segments, all
within the borders of the City of Toronto. The dataset included a number of descriptive
factors for each segment, including the AADT for each segment, the total number of
collisions on the segment, a breakdown of the types of collisions (fatal, injury, property
damage only). The collision counts for all of the arterial segments were collected over a

The intersection crash dataset covered the same time from 1998-2002, and consisted of
1888 intersections. 107 of these intersections did not have historical data for a full five
years, and were therefore excluded from analysis. The remaining intersections included
138 3-legged intersections and 1643 4-legged intersections. The intersections included
geographical factors relating to the intersection location, and breakdowns of the collision
type.

3.4.2 Microsimulation Model

The microsimulation model covered the waterfront area of the City of Toronto, extending
from Lake Ontario north to Dundas Street, and from Woodbine Avenue in the east to
Roncesvailles in the west. A map of the simulation area is shown in Figure 1. The
network consists of 4964 links and 1843 nodes, and represents the area in the early
2000’s. The network was built originally for the Toronto Waterfront Revitalization
Corporation, and more details regarding the model can be found in the Gardiner-Lake
Shore Scoping Study (2003). A screenshot of the simulation model is shown in Figure 2.
Figure 1: Map of the Study Area

Figure 2: Screenshot of Microsimulation Model
3.4.3 Study Area

The simulation network was an accurate depiction of the physical road network during the period 1998-2002. The arterial links were aggregated into segments that matched the arterial segments for which historical data exists. This resulted in 267 arterial segments represented in the simulation model. Additionally, 132 4-legged intersections from the historical data set were represented in the simulation model.

The arterial and intersections sets were divided into two sets, both randomly chosen. As a result, the calibration data set consisted of 134 arterials and 66 intersections. The second set for the prediction comparison had 133 arterial segments and 66 intersections.
4 Microsimulation Calibration

The calibration of the microsimulation model was initially attempted using three different seed values. The intent was to run each set of parameter values three times, and average the results. Unfortunately, SSAM analysis runs at much less than real-time, increasing with the resolution of the model. After a few sets of values were tested, it became apparent that multiple seeds would be too time consuming for calibration purposes. As a result, the calibration of the microsimulation model was performed using a seed value of 8651, and no other seed values.

4.1 Simulated Agent to Vehicle Ratio

Initial simulation runs were performed with one agent per 100 vehicles. Two sets of parameters were tested with agent to vehicle ratios (AVRs) of 1:50, 1:75, and 1:100. Paramics completed its simulation runs in 1-3 hours, running slower as more simulated agents were required. The output TRJ files increased from approximately 2 GB at a ratio of 1:100, to around 15-20 GB for simulations at a 1:50 ratio. Simulations were attempted at a ratio of 1:25, resulting in over 30 GB of data in the resultant TRJ files. However, SSAM would crash due to a lack of available memory for the analysis.

At the 1:50 agent to driver ratio, SSAM required over half a day to analyze the vehicular movements and produce a list of conflicts. As a result, simulation runs with a higher AVR were not examined. The computational time at this level was 3-4 times that at the 1:75 ratio, as was the input TRJ file size. A large portion of the increase in output conflicts is due to false conflicts at roadways that cross at different elevations.

The simulation network was set up with a mean headway of 0.75 seconds and a mean reaction time of 0.4 seconds for the first set of parameters, and a mean headway of 0.25 seconds and a mean headway of 0.3 seconds for the second set of parameters. The max TTC in both analyses was 1.5 seconds. The results of the different agent to driver ratios show that the performance of the collision prediction models improved as the resolution of the model increased, as was expected. The collision prediction models are outlined in Table 1. One item of note is that the models were fitted against 5 year crash counts, and
not yearly crash frequencies. In order to obtain an average yearly crash frequency, the model crash predictions would need to be divided by five. Dividing the 5 year crash counts would have resulted in non-integral average yearly crash counts, but the negative binomial model is based on a discrete distribution, so integral crash counts should be used when fitting the model.

Table 1: Agent to Vehicle Calibration CPMs

<table>
<thead>
<tr>
<th>Arterial (MHT = 0.75, MRT = 0.4)</th>
<th>Agent:Vehicle Ratio</th>
<th>1:50</th>
<th>1:75</th>
<th>1:100</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>0.11655141</td>
<td>0.077146111</td>
<td>0.051860541</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>-3.121185063</td>
<td>-3.197721726</td>
<td>-3.357788262</td>
<td></td>
</tr>
<tr>
<td>$SS_{err}$</td>
<td>110.7028856</td>
<td>114.5942314</td>
<td>114.9823085</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.041384975</td>
<td>0.00768845</td>
<td>0.004327955</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>46.86120678</td>
<td>48.73214875</td>
<td>48.78501388</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arterial (MHT = 0.25, MRT = 0.3)</th>
<th>Agent:Vehicle Ratio</th>
<th>1:50</th>
<th>1:75</th>
<th>1:100</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>0.136559162</td>
<td>0.088225268</td>
<td>0.078890047</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>-1.293974015</td>
<td>-3.102663537</td>
<td>-3.222400553</td>
<td></td>
</tr>
<tr>
<td>$SS_{err}$</td>
<td>112.8565515</td>
<td>114.0578349</td>
<td>114.2607192</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.02323616</td>
<td>0.012333295</td>
<td>0.010576449</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>47.76043224</td>
<td>48.46126873</td>
<td>48.75316865</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intersection (MHT = 0.75, MRT = 0.4)</th>
<th>Agent:Vehicle Ratio</th>
<th>1:50</th>
<th>1:75</th>
<th>1:100</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>0.284497193</td>
<td>0.321850817</td>
<td>0.365317702</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>-2.924746507</td>
<td>-2.70046854</td>
<td>-2.847295649</td>
<td></td>
</tr>
<tr>
<td>$SS_{err}$</td>
<td>27.18526833</td>
<td>29.97150999</td>
<td>25.92243635</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.271574279</td>
<td>0.196917297</td>
<td>0.305411697</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>12.51679</td>
<td>12.25650842</td>
<td>10.38701488</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intersection (MHT = 0.25, MRT = 0.3)</th>
<th>Agent:Vehicle Ratio</th>
<th>1:50</th>
<th>1:75</th>
<th>1:100</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>0.364561221</td>
<td>0.373175048</td>
<td>0.393680543</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>-2.442262338</td>
<td>-2.344806455</td>
<td>-2.588051447</td>
<td></td>
</tr>
<tr>
<td>$SS_{err}$</td>
<td>26.64491754</td>
<td>29.36060958</td>
<td>28.88871789</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.286052908</td>
<td>0.213286294</td>
<td>0.22593057</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>10.43119289</td>
<td>11.74565032</td>
<td>11.56451805</td>
<td></td>
</tr>
</tbody>
</table>

Model form: $\log(\text{Crashes}) = \alpha \times \log(\text{Conflicts}) - \beta$
The calibrated models show that the AVR has a significant effect on the model coefficients. In the case of arterials, the conflicts had an increasing explanatory role as the AVR decreased. The intercept decreased as the AVR became smaller, and more vehicles populated the model. Interestingly, the same effect was not noted for intersections. Conflicts decline in significance as an explanatory variable as the AVR was lowered. The R-square and chi-squared values were computed for these models in order to provide additional information.

For arterials, a lower AVR produces a better value for both R-square and chi-square. This assessment is consistent with the observation that conflicts are a better explanatory variable as the model resolution increases. In the case of intersections, the goodness of fit statistics vary in their assessment of AVR. The 1:75 agent to vehicle ratio resulted in the highest residual sum of squares and chi-squared value, and the lowest R-square value. The statistics show that with the first set of parameters (MHT = 0.75s, MRT = 0.4s), the 1:100 ratio performed better than the 1:50 ratio. With the second set of parameters (MHT = 0.25s, MRT = 0.3s), the 1:100 ratio performed better.

There is no easy explanation as to why the intersection models varied between which AVR led to better goodness of fit measurements. The explanatory power of conflicts was highest for both sets of parameters when the AVR was 1:100. However, this is very likely due to the fact that there were more conflicts resulting from the runs when the AVR was 1:50. With the second set of parameters, the intercept was closest to zero for the 1:50 AVR case. This is in agreement with the goodness of fit statistics indicating the 1:50 AVR was the better fitting model for those parameters. Given that the arterial models fared better when the AVR was set to 1:50, the chosen ratio for subsequent calibration and the prediction phase is 1:50.

The choice of a lower ratio implies a higher resolution simulation, and increases computational time. This limits the ability to test the effects of other parameters; however, as shown by the results of the intersection models, a set of parameters may lead to different conclusions with a different AVR. Given that a lower resolution model would
be less accurate in describing the traffic network, and was less useful in prediction arterial collisions, the lowest AVR was chosen for subsequent use.

4.2 Mean Headway Time and Reaction Time

The mean headway and mean reaction times were tested simultaneously. Sets of parameters values were defined, run, and compared to each other. The initial parameter values were MHT = 1.25 seconds and MRT = 0.61 seconds. The results using this parameter set were worse than sets with lower MHT and MRT values. Lower values for either parameter corresponds to more aggressive driving. Paramics suggests an MHT of 0.85-0.95 for urban areas; however, the MHT values that were tested were 0.25, 0.5, and 0.75 seconds. The default Paramics value for mean reaction time is 1.0 seconds. The model was previously calibrated with a 0.61 second reaction time; however, lower values were tested. The values used were 0.2, 0.3 and 0.4 seconds.

Each of the MHT values was used in a simulation run with each of the MRT values, with the exception of the MHT = 0.75s, MRT = 0.2s pair. By the point that set was to be tested, it had already been noted that an MHT value of 0.75 seconds was providing poorer quality predictions. In addition to the eight remaining possible combinations of MHT and MRT, two more parameter sets were proposed: MHT = 0.15s, MRT = 0.3s, and MHT 0.4s and MRT = 0.2s. Including the original parameters of MHT = 1.25s, MRT = 0.61s, that leaves simulation runs. The model parameters and the goodness of fit statistics for the resulting crash prediction models are shown in Table 2. As before, the crashes are five-year crash counts, not yearly crash rates.
Table 2: Driver Behaviour Calibration CPMs

<table>
<thead>
<tr>
<th>Arterial</th>
<th>MHT (s)</th>
<th>0.15</th>
<th>0.25</th>
<th>0.25</th>
<th>0.25</th>
<th>0.4</th>
<th>0.5</th>
<th>0.5</th>
<th>0.75</th>
<th>0.75</th>
<th>1.25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MRT (s)</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>α</td>
<td></td>
<td>0.1509</td>
<td>0.1734</td>
<td>0.1746</td>
<td>0.0947</td>
<td>0.2155</td>
<td>0.2083</td>
<td>0.1028</td>
<td>0.0989</td>
<td>0.1226</td>
<td>0.1166</td>
</tr>
<tr>
<td>SS&lt;sub&gt;err&lt;/sub&gt;</td>
<td></td>
<td>107.4263</td>
<td>108.6171</td>
<td>108.6183</td>
<td>113.9890</td>
<td>109.6423</td>
<td>109.4781</td>
<td>111.1407</td>
<td>112.7812</td>
<td>112.2759</td>
<td>110.7029</td>
</tr>
<tr>
<td>R²</td>
<td></td>
<td>0.0698</td>
<td>0.0594</td>
<td>0.0594</td>
<td>0.0129</td>
<td>0.0506</td>
<td>0.0520</td>
<td>0.0376</td>
<td>0.0239</td>
<td>0.0278</td>
<td>0.0414</td>
</tr>
<tr>
<td>χ²</td>
<td></td>
<td>44.9850</td>
<td>45.9112</td>
<td>45.1538</td>
<td>48.7651</td>
<td>46.4362</td>
<td>46.5799</td>
<td>47.1645</td>
<td>46.9704</td>
<td>47.3065</td>
<td>46.8612</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intersection</th>
<th>MHT (s)</th>
<th>0.15</th>
<th>0.25</th>
<th>0.25</th>
<th>0.25</th>
<th>0.4</th>
<th>0.5</th>
<th>0.5</th>
<th>0.75</th>
<th>0.75</th>
<th>1.25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MRT (s)</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>α</td>
<td></td>
<td>0.3168</td>
<td>0.3340</td>
<td>0.4100</td>
<td>0.3506</td>
<td>0.3649</td>
<td>0.3462</td>
<td>0.3248</td>
<td>0.2823</td>
<td>0.3017</td>
<td>0.2845</td>
</tr>
<tr>
<td>R²</td>
<td></td>
<td>0.2574</td>
<td>0.2060</td>
<td>0.3399</td>
<td>0.2836</td>
<td>0.2731</td>
<td>0.2622</td>
<td>0.3082</td>
<td>0.2206</td>
<td>0.2155</td>
<td>0.2716</td>
</tr>
</tbody>
</table>

Model form: log(Crashes) = α × log (Conflicts) − β
Both the intersection and arterial cases performed better with a lower MHT and MRT. Larger headway values generally lead to lower numbers of conflicts, since drivers are more spaced out and less likely to be placed in a conflict situation. Lower headway values are associated with more aggressive driving, and higher numbers of conflicts. Similarly, lower reaction times reflect driving which is more aggressive. The results indicate that the driving style in the simulation needs to be on the aggressive side for simulated conflicts to model crashes.

The MHT = 0.25s and MRT = 0.3s pair appears to have performed better than most. For the intersection case, it has the lowest residual sum of squares, highest R-square value, and lowest chi-squared value. Additionally, it has the lowest intercept value and the highest conflict coefficient, indicating that those two parameters provide conflicts with the best explanatory properties for crash prediction. For the arterial case, these parameters resulted in the second best residual sum of squares, R-square value, and chi-squared value.

The MHT = 0.4s and MRT = 0.25s parameter pair showed the highest coefficient value for conflicts in arterial collision prediction. The intercept value was the second lowest for arterials. The coefficients for the intersection case showed the second highest reliance on conflicts, with a conflict coefficient slightly below and an intercept slightly above the MHT = 0.25s and MRT = 0.3s parameter set. This parameter pair had the third best R-square value, fourth best residual sum of squares, and fifth best chi-squared value.

It is worth noting that the parameter that varied the most across the different models was the R-square. Especially in the case of arterials, R-square varied much more with different parameter values than either chi-squared or the residual sum of squares.

For the purposes of the last phase of this thesis, the chosen parameters were MHT = 0.4s and MRT = 0.2s. This decision was made due to the fact that those parameters performed well in all measures, and showed the highest use of conflicts as the explanatory variable for arterials, and third highest for intersections. Since the goal of this project is to evaluate the usefulness of conflicts for predicting crashes, the decision was made to go
with a set of parameters that would provide conflicts that would be most used as the explanatory variable in a collision prediction model.

### 4.3 Shortcomings

There are a number of other user definable parameters that could be explored. However, such an evaluation would have required significant amounts of computational time to process. For this reason, other factors such as minimum gap and maximum time to collision were not examined. The default value for minimum gap in Paramics is 2 metres. A decrease in the minimum gap length would likely cause an increase in conflicts as vehicles travel closer together. The most likely result would be a worse performance in arterial predictions, as vehicles get too close to each other when decelerating to approach an intersection. Rear-end conflicts would occur due to vehicles not stopping quickly enough, which would be counted as an arterial conflict since both vehicles are on the same link. However, such an accident in real life may be considered an intersection collision, depending on how far from the intersection the collision occurs.

The number of conflicts that occurs during a simulation run is affected by the definition of what constitutes a conflict, which is governed by a predefined time to collision (TTC). For the purposes of this study, a threshold TTC of 1.5 seconds was selected. This value is fairly commonly encountered in conflict analyses; however, lower TTC values are also often used. A lower TTC would provide conflicts of a higher severity, and it is possible that these more severe collisions would provide a more accurate depiction of reported accidents; since it is serious incidents which are most reported. The longer TTC conflicts may correlate to minor accidents, which are infrequently reported. Since those crashes are unreported, it is not possible to test if this hypothesis is true.

On a final note regarding microsimulation calibration, the factors adjusted affect simulated driver behaviour. Other parameters, such as familiarity and perturbation, affect routing much more than driver behaviour. These factors may have affected the conflict output, but they are also more likely to change the traffic volumes in the network. Such a change would be reflected in both the conflict-based predictions, as well as the volume-based predictions. If these types of parameters were to be tested in the future, both
volume and conflict based predictions should be used as a benchmark in the calibration process. A more thorough calibration process would have also looked at the more traditional outputs, such as traffic counts and queue lengths. Once again, this would have been considered were more computational resources available at the time of data processing. Cursory glances at the traffic counts were taken to ensure they roughly matched up with the arterial and intersection AADT, but no steps were taken to factor that in when choosing parameter values.
5 Model Fitting

Models are needed in order to generate crash predictions. The calibration of the models occurred with half of the arterials and intersection in the model. The other half was reserved to generate predictions. The simulation model was run eight times and the average conflict counts and traffic volumes were calculated. These averages form the dataset from which both volume-based and conflict-based models were calibrated and from which predictions were generated.

One volume-based model and four conflict-based models are used in the generation of crash predictions. The volume-based models are negative binomial regression models. For the conflict-based models, two linearized models and two negative binomial models are prepared. Two of the models are rescaled versions of the SSAM, and do not need to be fitted. The other two conflict-based models, one of each type, need to be fitted.

For the volume-based model, two models were fitted. Originally, models were fitted using historical data from the city of Toronto, but outside of the model area. The arterial and intersection models fitted using city-wide data were compared with models fitted using the simulation network set of arterials and intersections that was used to fit the conflict-based model, in order to determine what differences would arise.

5.1 Arterial Models

The linearized and non-linearized conflict-based model forms were fitted using R. The coefficients of both models as well as the goodness of fit for the models are presented in Table 3. All coefficients were statistically significant at the 0.01 level of significance. The models were fitted using 134 arterial segments, and both models had 133 degrees of freedom.
Table 3: Fitted Arterial Conflict-Based CPMs

<table>
<thead>
<tr>
<th>Model form</th>
<th>Crash = Conflicts^α × e^β</th>
<th>log(crashes) = α × log(conflicts) + β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>α: 0.30202 (0.06929)</td>
<td>α: 0.35497 (0.08032)</td>
</tr>
<tr>
<td></td>
<td>β: 2.57186 (0.31056)</td>
<td>β: 1.96330 (0.35922)</td>
</tr>
<tr>
<td>χ²</td>
<td>151.1697</td>
<td>121.5321</td>
</tr>
<tr>
<td>R² / R²</td>
<td>0.1093342</td>
<td>0.08395276</td>
</tr>
<tr>
<td>α</td>
<td>0.645926</td>
<td>N/A</td>
</tr>
<tr>
<td>Scaled deviance</td>
<td>1.113686</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*Standard error in parentheses next to coefficient values*

The negative binomial model form had a worse chi-squared value when compared to the linear model; however, the chi-squared value for the linearized model is calculated using residuals based on the predicted variable being the logarithm of crash counts. Correcting for this by taking the exponential of the linearized model residuals produces a chi-squared value of 511.224. Neither model is a great fit for the collision data, based on the R-square and R² values calculated, as well as the chi-squared valued, the negative binomial model was the better of the two.

The arterial volume-based model fitted with city-wide data is shown in Equation 12, and the model fitted with the simulation network arterial set is shown in Equation 13. L represents the length of the arterial segment. The number of lanes is self-explanatory, and one way is a Boolean variable indicating if the segment is one way only. The three access variables are scores used to denote how many laneways, roadways, and driveways open onto the arterial segment.
\[ \text{Crashes} = L^{a_1} \times AADT^{a_2} \times \exp(\beta_0) \]
\[ + \text{number of lanes} \times \beta_1 + \text{one way} \times \beta_2 \]
\[ + SP \times \beta_3 + SS \times \beta_4 \]
\[ + \text{laneway access} \times \beta_5 \]
\[ + \text{roadway access} \times \beta_6 \]
\[ + \text{driveway access} \times \beta_7 \] \]  
Equation 12

\[ \text{Crashes} = L^{a_1} \times AADT^{a_2} \times \exp(\beta) \]  
Equation 13

The coefficients of both models are detailed in Table 4. All coefficients were statistically significant at the 0.01 level of significance, or better. The arterial model was fitted with 1039 data points, and 743 were used to create the intersection model. The chi-squared value was scaled by the degrees of freedom to provide a simple means of comparison.

**Table 4: Fitted Arterial Volume-Based CPMs**

<table>
<thead>
<tr>
<th>Model Fit Set</th>
<th>City-wide</th>
<th>Simulation model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>-$\alpha_1$: 0.227189 (0.038044) &amp; $\alpha_1$: 0.92580 (0.07677)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-$\alpha_2$: 0.638083 (0.048091) &amp; $\alpha_2$: 0.18145 (0.08584)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_0$: -5.295525 (0.526707) &amp; $\beta$: -2.81909 (0.87985)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_1$: 0.093026 (0.022573) &amp;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_2$: -0.608340 (0.197038) &amp;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_3$: 0.557770 (0.061679) &amp;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_4$: 0.676001 (0.161534) &amp;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_5$: 0.144108 (0.029443) &amp;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_6$: 0.084817 (0.008211) &amp;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\beta_7$: 0.031822 (0.003190) &amp;</td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>1196.227 &amp; 128.41567</td>
<td></td>
</tr>
<tr>
<td>Scaled $\chi^2$</td>
<td>1.162514 &amp; 0.972846</td>
<td></td>
</tr>
<tr>
<td>$R^2_a$</td>
<td>0.401644 &amp; 0.5178764</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.350448 &amp; 0.3450969</td>
<td></td>
</tr>
<tr>
<td>Scaled deviance</td>
<td>1.082847 &amp; 1.079899</td>
<td></td>
</tr>
</tbody>
</table>

*Standard error in parentheses next to coefficient values*
The two volume-based models have a different number of explanatory variables due to the fact that many of the variables that were significant when calibrating to the larger city dataset were no longer significant when attempting to calibrate the with the smaller data set from the simulation model. Were the simulation network larger, it is possible that some of the removed explanatory variables would once again be significant.

The first model fitted with city wide data is a much poorer model. The R-alpha value increased from 0.40 to 0.51 when the model was fitted only against the simulation data. The scaled chi-squared decreased from 1.16 to 0.97 when the model was fitted using only the simulation network area. Both the chi square and $R^2$ values indicate that the model calibrated to the simulation data set is a better fit for the data it was calibrated to.

A possible cause for the wide difference in goodness of fit is different driving behaviour in the different parts of the city represented in the city-wide dataset. The city-wide data includes arterial segments that are located both downtown and in the suburbs of the city. In contrast, the simulation network only covers the downtown core of the city. It is very likely that drivers behave differently when driving downtown compared to suburban driving. While other explanatory variables could have accounted for the different driving style in different parts of the city, such as the number of lanes on the arterial segment, these values were statistically insignificant when included in the model.

When comparing between conflicts and volume as an explanatory variable, it appears that conflicts are poor when fitting prediction models. Based on the significance of segment length as a factor in the volume-based models, a new conflict-based prediction model including length was formed, as shown in Equation 14. The coefficients and goodness of fit measures are described in Table 5. All coefficients were significant at the 0.01 level.

$$Crashes = L^{a_1} \times Conflicts^{a_2} \times \exp(\beta)$$  
Equation 14
Table 5: Revised Arterial CPM

<table>
<thead>
<tr>
<th>Model form</th>
<th>Crashes = Conflicts^α1×eβ1</th>
<th>Crashes = L^α2×Conflicts^α3×exp(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>0.30202 (0.06929)</td>
<td>0.88200 (0.07691)</td>
</tr>
<tr>
<td>β</td>
<td>2.57186 (0.31056)</td>
<td>1.73503 (0.42446)</td>
</tr>
<tr>
<td>χ²</td>
<td>151.1697</td>
<td>130.6045</td>
</tr>
<tr>
<td>R²</td>
<td>0.1093342</td>
<td>0.5372259</td>
</tr>
<tr>
<td>α</td>
<td>0.645926</td>
<td>0.3312469</td>
</tr>
<tr>
<td>β</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scaled deviance</td>
<td>1.113686</td>
<td>1.072124</td>
</tr>
</tbody>
</table>

Standard error in parentheses next to coefficient values

With the included length term, the model performance improves greatly. The chi-squared value of the revised model is 130, only slightly worse than the value of 128 obtained by the volume-based version. The $R^2_a$ value for the revised conflict-based model is 0.53, better than the 0.51 from the volume-based model. The scaled deviance for both models is also very similar, at 1.072 for the revised conflict-based model compared to 1.080 for the volume-based model. In short, both conflict and volume based crash prediction models can be fit for arterial roadway, but the question remains how their predictions perform.

5.2 Intersection Models

The conflict based models, both linearized and non-linearized, were fitted using R and are described in Table 6. The coefficients were all statistically significant at the 0.01 level or better. The models were fit using 64 intersections, resulting in each model having 63 degrees of freedom.

Table 6: Fitted Intersection Conflict-Based CPMs

<table>
<thead>
<tr>
<th>Model form</th>
<th>Crashes = Conflicts^α1×eβ1</th>
<th>log(crashes) = α×log(conflicts) + β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>0.38921 (0.07405)</td>
<td>0.40707 (0.07455)</td>
</tr>
<tr>
<td>β</td>
<td>2.60987 (0.30142)</td>
<td>2.36639 (0.30212)</td>
</tr>
<tr>
<td>Pearson chi-square</td>
<td>85.62785</td>
<td>21.53934</td>
</tr>
<tr>
<td>$R^2_a$ / R-square</td>
<td>0.2461375</td>
<td>0.1627950</td>
</tr>
<tr>
<td>α</td>
<td>0.3113761</td>
<td>N/A</td>
</tr>
<tr>
<td>Scaled deviance</td>
<td>1.075423</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Standard error in parentheses next to coefficient values
The negative binomial model form, upon first glance, had a worse chi-squared value when compared to the linear model. However, as with the arterial model previously, the chi-squared value needs to be adjusted to use residuals based on crash counts, and not their logarithm. Taking this adjustment into account produces a chi-squared value of 131.6183. The R-square value for the linearized model indicates a poor fit, but the $R^2_a$ value for the negative binomial model is better than the linearized model. Once again, the negative binomial model provides a better fit.

The volume-based model fitted with city-wide data is shown in Equation 15, and the model fitted with simulated data is shown in Equation 16. CPV is a Boolean value representing pedestrian volumes more than 1000 per day. This factor remained significant at the 0.01 level as did the AADT values for the model fitted with city-wide data. The coefficients of both models and their goodness of fit statistics are listed in Table 7.

\[
Crashes = AADT_m^{a_1} \times AADT_{min}^{a_2} \times \exp (\beta_0 + CPV \times \beta_1) \quad \text{Equation 15}
\]

\[
Crashes = AADT_{ma}^{a_1} \times AADT_{min}^{a_2} \times \exp (\beta) \quad \text{Equation 16}
\]

<table>
<thead>
<tr>
<th>Table 7: Fitted Intersection Volume-Based CPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Fit Set</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Coefficients</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
</tr>
<tr>
<td>Scaled $\chi^2$</td>
</tr>
<tr>
<td>$R^2_a$</td>
</tr>
<tr>
<td>$\alpha$</td>
</tr>
<tr>
<td>Scaled deviance</td>
</tr>
</tbody>
</table>

*Standard error in parentheses next to coefficient values*

When fitting the model with the simulation network data, the $R^2_a$ value for intersection models decreased from 0.71 to 0.12, and the scaled chi-squared value increased from 0.96 to 1.29. This difference in fit between crash predictions models fitted on historical data
and simulated data was not expected to be so great at intersections. It appears that having fewer data points to fit the intersection model severely decreased its fitting capability. Indeed, this is shown by the fact that only $AADT_{min}$ proved to be significant, and only at the 0.05 level of significance. Both the intercept and the $AADT_{maj}$ were statistically insignificant in the model.

Comparing the conflict and volume based models fitted on the same data set shows that the conflict-based model performed better. The chi-squared value of the conflict-based model is higher (85.6 compared to 80.1), but the $R^2$ of the conflict-based models is much better (0.24 compared to 0.12). More importantly, all coefficients in the conflict-based model are statistically significant at the 0.01 level, something that is not true for all of the coefficients in the volume-based model.

The failure to fit a volume-based model to the simulated intersection data set was unexpected. It is possible that a fit could not be achieved due a low number of intersections being available to calibrate the model. However, a conflict-based negative binomial was fitted using the same series of data. Since the conflict-based model has only two coefficients to fit, compared to three or more for an intersection volume-based model, fewer data points are needed to calibrate a conflict-based model. This indicates the potential of conflict-based collision prediction models to be feasible in cases where there are not many data points to calibrate a volume-based model.

For the purposes of generating predictions to test, both the city-wide and simulation network calibrated intersection volume-based CPMs were used. While it would be simple to expect the city-wide fitted model to generate better fitting predictions, it is possible that this will not be the case. If the driving behaviour at downtown intersections varies from suburban behaviour, the city-wide calibration may have mis-calibrated the model for downtown driving. As a result, both models were used to generate predictions with no expectation as to which one will provide better predictions.
6 Collision Predictions

The main objective of this thesis was to test the predictive abilities of conflict and volume based models, and compare the predictions generated with the two models in comparison with historical crash counts. Now that collision prediction models have been calibrated, it was then possible to generate predictions to evaluate. Comparing these models allowed a fair assessment of how well the models could predict collisions, as opposed to the typical approach with volume-based models to explain what factors were important.

The model forms calibrated in the previous chapter were used to generate predictions. In addition to the two arterial conflict-based model and the three intersection conflict-based models, rescaled versions of the two SSAM models were tested for both arterials and intersections. Two models were rescaled versions of the SSAM models (Equation 17 and Equation 18), and the other two were calibrated from scratch. The rescaling of the two SSAM model forms and calibration of the two other models was done for both arterials and intersections.

\[
\log(\text{Crashes}) = 1.09 \times \log(\text{Conflicts}) - 0.98 \quad \text{Equation 17}
\]

\[
\text{Crashes} = 0.119 \times \text{Conflicts}^{1.419} \quad \text{Equation 18}
\]

In order to rescale the SSAM models, one parameter in each model was adjusted so that the sum of predicted crashes equalled the sum of historical crashes. The factors that were rescaled are the intercept in the first model (0.98 in Equation 17), and the scale coefficient in the second model (0.119 in Equation 18). This rescaling allowed the testing of the transferability of the SSAM model, and accounted for the fact that the historical data is for a five-year period. The predictions generated were all over a five year period, for the sake of consistency.

The simulated flows generated by Paramics were examined prior to prediction generation, and any locations where the simulated volumes in either the major or minor directions varied by a factor of 5 or more were excluded from the analysis. This resulted in 133 arterial segments and 64 intersections to use for prediction generation. Further filtering could be done, however this would have removed many more data points.
Additionally, the objective was to test the predictive ability of the models when based on the same data. Some of the errors were due to the difference between the simulated traffic in the network and historical traffic, while the remainder was due to the error caused by the models. Since there was no historical conflict data to compare simulated conflicts with, it would be unfair to pick only the data points which would generate similar predictions as those based only on historical traffic flows.

6.1 Arterial Models

SSAM rescaling provided the two models in Equation 19 and Equation 20. The original coefficients for the linearized and non-linearized models led to a total of 9684 and 17545 crashes, respectively. The total number of arterial conflicts that historically occurred was 5651, or 1130 per year. Since the SSAM models were over-predicting crashes, the microsimulation model being used must be generating a higher number of conflicts per hour than the model used in the SSAM validation.

\[
\log(\text{Crashes}) = 1.09 \times \log(\text{Conflicts}) - 1.519 \quad \text{Equation 19}
\]

\[
\text{Crashes} = 0.0383 \times \text{Conflicts}^{1.419} \quad \text{Equation 20}
\]

Predictions were generated using data from the 133 arterial segments not used in the fitting of each of the models. The chi-squared and R-square values computed between the historical crash counts and the predicted counts are shown in Table 8.

<table>
<thead>
<tr>
<th>Model form</th>
<th>(\chi^2)</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{Crashes} = L^{a_1} \times AADT^{a_2} \times \exp(\beta))</td>
<td>46.00625</td>
<td>0.141624</td>
</tr>
<tr>
<td>(\log(\text{Crashes}) = 1.09 \times \log(\text{Conflicts}) + \beta)</td>
<td>102.6652</td>
<td>-1.09422</td>
</tr>
<tr>
<td>(\text{Crashes} = a \times \text{Conflicts}^{1.419})</td>
<td>133.6178</td>
<td>-2.26664</td>
</tr>
<tr>
<td>(\log(\text{Crashes}) = a \times \log(\text{Conflicts}) + \beta)</td>
<td>53.08109</td>
<td>0.015165</td>
</tr>
<tr>
<td>(\text{Crashes} = \text{Conflicts}^a \times \exp(\beta))</td>
<td>108.9541</td>
<td>-0.04547</td>
</tr>
<tr>
<td>(\text{Crashes} = L^{a_1} \times \text{Conflicts}^{a_2} \times \exp(\beta))</td>
<td>51.01834</td>
<td>0.039964</td>
</tr>
</tbody>
</table>

The arterial volume-based model shows reasonable performance. The CURE plot is shown in Figure 3. The cumulative residuals oscillate about zero, and stay within two
standard deviations. The chi-squared value is the lowest of the models, and the R-square value is the highest of the group. Without a doubt this is the best of the predictive models.

**Figure 3: Arterial CURE Plot for Volume-Based CPM**

Both the SSAM linearized and non-linearized models performed poorly. The CURE plot for the linearized model is shown in Figure 4 and non-linear model in Figure 5. In both cases the cumulative residuals are generally greater than two standard deviations away from the baseline. The chi-squared and R-square of the non-linearized model was the worst of all of the models, and the linearized version the third worst. In both cases the chi-squared value is more than double that of the volume-based model, and the R-square values were both negative.
Figure 4: Arterial CURE Plot for SSAM Linearized CPM

For the case of the linear conflict-based arterial model, both the chi-squared and R-square values were the third best. When examining the CURE plot (Figure 6), the model fit is worse than the volume-based CPM. At times the cumulative residuals exceed two standard deviations, but for the most part are within that bound. For lower values of
conflicts, this conflict model trends towards over-predicting crashes. The only oscillations about zero occur with higher numbers of conflicts (> 200).

Figure 6: Arterial CURE Plot for Linear Conflict-Based CPM

![Figure 6: Arterial CURE Plot for Linear Conflict-Based CPM](image)

Figure 7: Arterial CURE Plot for NB Conflict-Based CPM

![Figure 7: Arterial CURE Plot for NB Conflict-Based CPM](image)
The CURE plot above (Figure 7) is for the conflict-based negative binomial model. Judging by the CURE plot alone, it does not appear too bad. The cumulative residuals stay within two standard deviations, and the CURE plot is the best of the conflict-based models. However the other goodness of fit measures were poor: the chi-squared value was the second worse, and the R-square was the third worst. If it were not for the poor chi-squared value, the negative binomial model would have undoubtedly been the best of the conflict-based models.

When the length factor was added to the negative binomial conflict-based model, the goodness of fit statistics improved greatly. The length and conflict-based model had the second lowest chi-squared value, and second highest R-square value. Evaluating the CURE plot (Figure 8) shows that the cumulative residuals do exceed two standard deviations for conflict numbers approximately between 80 and 180. Additionally, there is very little oscillation in the cumulative residual line about zero. The CURE plot for the negative binomial conflict-based model without length is better than the model including length.

Figure 8: Arterial CURE Plot for NB Length and Conflict-Based CPM

The arterial conflict-based models all performed poorly compared to the volume-based collision predictions. The poor performance of the SSAM models shows that conflict-
based models do need to be fitted specifically for the network they are being used on, at least for arterial segments. When looking at fitted conflict-based models, there is a trend of over-predicting crash counts at lower values of conflicts. At higher values of conflicts, the conflict-models are more likely to oscillate about zero, indicating a better fit when conflicts counts are higher.

The performance of the intersection models is expected to be better, given the conflict-based models fit better during the model calibration phase. If the conflict-prediction models do in fact work better for intersection cases, some explanation was required as to why conflict-based prediction performs much worse at arterials than at intersections. The volume-based predictions will also need to be compared to determine if there is an issue related to the routing of the simulated vehicles or instead a specific problem when generating arterial conflicts. This will be discussed after the intersection results are presented, in chapter 6.3.

### 6.2 Intersection Models

The rescaling of the SSAM models for the intersection data provides the coefficients shown in Equation 21 and Equation 22. The original coefficients for the linearized and non-linearized models led to a total of 3118 and 4888 crashes, respectively. The total number of arterial conflicts that historically occurred was 3937, or 787 per year. As was the case with the arterial set, the SSAM model predictions were much higher than historically occurred.

\[ \log(Crashes) = 1.09 \times \log(Conflicts) - 0.747 \]  
Equation 21

\[ Crashes = 0.0958 \times Conflicts^{1.419} \]  
Equation 22

The SSAM models were recalibrated so that the mean prediction from the models was the same as the mean observed historical crash frequency. The recalibrated models provided poor results. The chi-squared and R-square values for the SSAM models, as well as the other models tested are listed in Table 9.
Table 9: Intersection Crash Prediction Goodness of Fit

<table>
<thead>
<tr>
<th>Model form</th>
<th>$\chi^2$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Crashes = AADT_{maj}^{a_1} \times AADT_{min}^{a_2} \times \exp (\beta)$</td>
<td>35.40546</td>
<td>0.253713</td>
</tr>
<tr>
<td>$Crashes = AADT_m^{a_1} \times AADT_{min}^{a_2} \times \exp (\beta_0)$</td>
<td>19.6358</td>
<td>-0.21446</td>
</tr>
<tr>
<td>$\log(Crashes) = 1.09 \times \log(Conflicts) + \beta$</td>
<td>35.27008</td>
<td>-0.63734</td>
</tr>
<tr>
<td>$\log(Crashes) = \alpha \times Conflicts^{1.419}$</td>
<td>53.39407</td>
<td>-1.49066</td>
</tr>
<tr>
<td>$\log(Crashes) = \alpha \times \log(Conflicts) + \beta$</td>
<td>18.42486</td>
<td>0.25607</td>
</tr>
<tr>
<td>$Crashes = Conflicts^{a_1} \times \exp (\beta)$</td>
<td>27.08878</td>
<td>0.194733</td>
</tr>
</tbody>
</table>

The goodness of fit statistics suggest that the model-fitted volume-based model was not the best model. The chi-squared value was the second worst, while the R-squared value was the second best. When looking at the major AADT CURE plot (Figure 9), the model does not appear to be well-behaved. The cumulative residuals remain within two standard deviations most of the time, but there is no oscillation about zero for AADT values under 8000 vehicles per day. The minor AADT CURE plot (Figure 10) is worse than the major AADT CURE plot, with the cumulative residuals near or exceeding two standard deviations most of the time. One data point was removed from the CURE plot, an outlier with a major AADT of over 21,000 vehicles per day. This data point remained for the goodness of fit statistics.
The volume-based model calibrated to city-wide data had a good chi-squared value, but a poor R-square value. The major AADT CURE plot for this model is shown in Figure 11. Based on the cumulative residuals, this model performed very poorly with respect to
major AADT. Compared to the major AADT CURE plot for the simulation-fitted model, this model is worse. The CURE plot based on the minor AADT (Figure 12) is better, but still far from ideal. As with the other volume-based model, one data point was removed from the CURE plot, an outlier with a major AADT of over 21,000 vehicles per day. This data point remained for the goodness of fit statistics.

**Figure 11: Major AADT Intersection CURE Plot for City-Fitted Volume-Based CPM**
The two rescaled SSAM models had the worst R-square value, and the worst and third worst chi-squared values. The linearized model had a chi-squared value only marginally better (35.27 compared to 35.40) than the volume-based model. The CURE plot for the linearized model is shown in Figure 13, and the non-linearized model in Figure 14. Both CURE plots show a similar picture. In both cases, the cumulative residuals are for the most part greater than two standard deviations. On the basis of the CURE plots and the goodness of fit measures, the SSAM recalibration failed.
The linear conflict-based model showed the best chi-squared and the best R-square value of the model predictions. The CURE plot is shown in Figure 15, and it is better behaved than any of the previous three models. The cumulative residuals always remain within
two standard deviations, and the cumulative residuals oscillate about zero a number of times.

**Figure 15: Intersection CURE Plot for Linear Conflict-Based CPM**

The last model used to generate predictions was the negative binomial conflict-based model. This model had the second best chi-square, and the third best R-square value. The CURE plot is shown in Figure 16, and it appears well behaved. The cumulative residuals stay within two standard deviations, and there are oscillations about zero. There are more oscillations about zero for this model than the non-linearized conflict-based model, and the cumulative residuals appear to be closer to zero than the non-linearized model. This is verified by graphing the cumulative residuals for both models on the same graph, as shown in Figure 17.
Although the comparison of CURE plots could lead to the conclusion that the negative binomial prediction model performed better, the better chi-squared value indicates that the total sum of squares error is lower for the linearized model. SSAM validation found the non-linearized conflict-based model to have a better fit based on goodness of fit.
measures such as chi-squared and R-square, however, those were based on model fitting. The fitting of the two models showed the negative binomial model had the worse chi-squared value, so it is not unreasonable that the negative binomial predictions have a higher chi-squared value than the linear conflict-based predictions.

Both conflict-based models had a better predictive performance than the volume-based models. Based on the poor fit of the volume-based model fitted using the simulation area data, this was not entirely unexpected. Using only goodness of fit statistics such as R-square and chi-square, it might not be expected that the volume-based model performed poorly, given the high R-square value of the first volume-based model; however, the CURE plots indicate clearly that the conflict-based models provide a better fit to historical collision data.

6.3 Discussion

The predictions generated indicate that conflict-based models provide better predictions at intersections, and worse predictions at arterials. This matches up with the observation that when fitted using the same data set, conflict-based models have a better fit than volume-based models at intersections, and the opposite holds true for arterials. When comparing the coefficient for conflicts between the intersection and arterial models, the coefficient is higher for the arterial models, indicating that fewer conflicts are being used to generate collision predictions.

Considering that more conflicts at intersections lead to better predictions, one possible cause of the poor performance of conflict-based models may be poor conflict generation on arterials. Paramics is built on car-following and lane-change models, which governs driver behaviour on arterials. These links are simulated in only one dimension, using only the vehicle’s distance, speed, and acceleration to determine driver actions. In contrast, vehicular movements at intersections are modelled much more elaborately (Duncan, 1997). The fact that movements at intersections are much more calculated likely leads to more accurate vehicular movement, resulting in conflicts which are a better representation of those that occur in real life.
The car-following model in Paramics mainly affects the speed that vehicles travel at. This includes deciding when to brake and when to accelerate. The other driver governing model is a lane-change model. It is responsible for the deciding when and if a simulated driver will change lanes. While observing the simulated traffic, it was noticed that vehicles usually change lanes when it is necessary. This is not always the case in real life, as there are many situations where lane changes occur which would not incite a lane change by simulated agents.

In Paramics, lane changes occur when a simulated agent has a desire to change lanes, and there is sufficient space to do so. A driver will desire to change lanes when there is a predictable hindrance to that driver’s movement in its current lane. Examples of such obstacles would include lane endings, stopped vehicles, and slow moving vehicles. Lane changes are also required if the next turn must be performed from a different lane than the vehicle is currently in. Generally, simulated drivers will not make lots of unnecessary lane change maneuvers. In reality, many drivers switch lanes much more often, and much more than necessary.

Reasons for lane changes in real life are much more diverse than merely maneuvering to make a turn. In many cases, drivers fail to anticipate a lane change that is required in order to perform a turn. This can result in drivers turning from the incorrect lane, however, much more frequently, this will result in a last minute lane change into the turning lane, and often when there is not a sufficient gap for the vehicle. These kinds of lane changes will not be modeled by simulated drivers since they were in the turning lane in advance, and not force their way into a lane. Additional, no driving agents were observed stopping in traffic in order to perform a lane change.

Another case where simulation does not reflect reality is with drivers who weave in and out of traffic. Within simulation there are no agents who drive at aggressively high speeds, and change lanes very frequently, in order to avoid slowing down. This behaviour is more relevant to freeway segments, but there are arterial cases where intersections are far apart where this behaviour may be noticed. This behaviour could be programmed by modifying a percentage of simulated agents to speed above the speed limit, and changing
their behaviour by prioritizing lane changes over deceleration so that these modified
drivers weave through traffic. Such a behavioural change would need to be calibrated of
course, potentially using records of the number of speeding infractions and other traffic-
related offences.

Simulated agents always are aware what lane they need to be in for upcoming turns. This
is something that can affect intersection behaviour, where a driver may choose to turn
from the incorrect lane, but more commonly will manifest itself in drivers who slowdown
in order to find a gap, and if a suitable gap is not found, forcing their way into the lane
they need to be in. Drivers do not always plan ahead in real life, or sometimes get
distracted, and this sort of delayed knowledge could possibly be simulated by altering the
routing logic used by simulation agents. Implementation would require the ability for
simulated agents to “zone out” and simply drive straight ahead, without regard to where
they are. Only when simulated agents are within a threshold distance from where their
next turn is will they become aware enough and then attempt to manoeuvre into position
to make any necessary turns. Of course, this then requires the programming to consider
what to do in the event a minimum gap to change lanes is not found.

When drivers are faced with the dilemma of forcing a lane change or changing route,
more aggressive drivers will tend towards making their way into a queue in an adjacent
lane. Some drivers will slow down and completely stop until someone permits enough of
a gap to allow a driver in. On the other hand, some drivers will simply forgo the lane
change and change their route. Simulating passive drivers is easier since it merely
requires instructing the simulated agent to change their route in the event they cannot
make a lane change in time. Simulating an agent forcing their way into a different lane is
a much harder thing to program. This behaviour requires the agent to partially perform a
lane change into a space smaller than their vehicle; if the minimum gap was simply
changed to be smaller than vehicle size two simulated vehicles could end up occupying
the same space in simulation. A new vehicle manoeuvre needs to be defined where a
simulated agent will obstruct their current lane and infringe on the adjacent lane until a
lane change occurs. These motions would only be undertaken when the lane change is
urgent.
An urgency factor, such as distance until a lane change must be made. In the case of turning at an intersection, this would be the distance to the intersection. Drivers would need a threshold distance before which a normal lane change will be attempted, and after which the simulated agent will force a lane change. This urgency factor will allow a forced lane change to be implemented in other cases. For example, merging behaviour caused by a lane reduction may include some cases where aggressive drivers force their way into the adjacent lane.

Forced lane changes are a situation that occurs only when the target lane is filled with a queue. In order to ensure queues are formed, the microsimulation model must be run with many simulated agents, therefore causing an increase in the number of conflicts. It is possible that simply running the simulation network with a lower simulated agent to driver ratio will improve the performance of the conflicts generated.

In order for simulated conflicts to be representative of conflicts created by real world drivers, the driving behaviour of the simulated agents needs to be as realistic as possible. Simulated agents will never fully replicate human driving behaviour, but any efforts to increase their accuracy can only improve the generated of simulated conflicts. These simulated conflicts will then be able to provide a better fit to crashes and better predictions than at this point in time. Improvements to the lane changing behaviour are the most obvious flaws that can be corrected with simulated driver behaviour. Adding the ability to force lane changes, weave in and out of traffic, and lose concentration are three potential additions that could be used to improve the realism of simulated drivers, and generate more realistic conflicts.
7 Conclusions

Conflicts can be used to make collision predictions in some cases. The conflict-based prediction models work well at intersections, but perform poorly at arterials. A possible explanation is that driving behaviour on arterial segments is not as detailed compared to the vehicular movements at intersections. Improvements to driver behaviour models, specifically the lane changing model, could be made to improve the performance of arterial conflict-based collision prediction models. The objectives of this thesis were to test the predictive ability of conflict-based collision-prediction models, and test their suitability for arterial segments. The transferability of the SSAM models was also to be examined, to determine if conflict-based models need to be calibrated for each test area.

The first step of this research was the calibration of the microsimulation model. The objective was to minimize the error between the historical crash counts and the predictions made by a simple conflict-based crash prediction model. Low values for mean headway time and mean reaction time were found to work best, causing aggressive behaviour by the simulated drivers. It was also determined that the number of simulated agents in the model has a large impact on the fit of crash predictions generated from conflicts. Unfortunately, an optimal ratio could not be determined since performance increased as the ratio decreased, and low ratios caused the simulation model to fail.

After the microsimulation model was calibrated, collision prediction models were calibrated. For arterials, the volume-based model had the best fit. The fit was best when estimated using only the simulation model area data, likely due to over fitting when using the city-wide data set. The poor performance of the conflict-based models led to the introduction of a length and conflict-based model, which fared better and fit comparably to the volume-based model. In the case of intersections, the volume-based model fit best when based on city-wide data. When fitted only with data from the simulation area, the volume-based model performed poorly. The conflict-based models both fit poorly, but not as bad as the volume-based model fitted to the simulation area data.

The predictions generated using the models fitted show that volume-based predictions outperform conflict-based predictions for arterials. Intersection predictions show that
conflict-based models can perform at least as good as the volume-based ones, regardless of which data it was fitted to. In both cases, the SSAM models performed terribly, and it is evident that conflict-based models are not transferable. Possible causes for the discrepancy between the good performance of the conflict-based model at intersections, and the poor performance at arterials, were discussed. Improvements are proposed for altering the lane changing model used in Paramics with the aim of improving the performance of the arterial conflict-based model.

Through the work in this thesis, it is evident that work must be done before conflicts can be used to predict crash frequencies for arterials. The conflict-based crash prediction models for intersection do perform well, indicating that arterial conflicts are less robust than those at intersections. The next step for simulation-based crash-prediction models is the creation of improved the agent vehicular movements while cruising, in order to improve the generation of arterial conflicts. Additionally, public transit could be added into the simulation to test if simulated conflicts with transit vehicles also can predict crashes.

Conflict-based models provide an alternative to volume-based models, and allow a tool for evaluating proposed projects. As traffic continues to increase, and road collisions account for an increasing percentage of fatalities, road safety will become a higher priority when evaluating infrastructure projects.
References


Gardner-Lake Shore Scoping Study (2002-2003), IntelliCAN Transportation Systems Inc. for the Toronto Waterfront Revitalization Corporation


