In compliance with the
Canadian Privacy Legislation
some supporting forms
may have been removed from
this dissertation.

While these forms may be included
in the document page count,
their removal does not represent
any loss of content from the dissertation.

Development of a GIS-Based Safety Information System for Rural Highways

Submitted by
Ossama E. M. Ramadan, B. Eng.

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment
of the requirements for the degree of Master of Applied Science

Department of Civil and Environmental Engineering
Carleton University
Ottawa, Ontario

© Ossama E. M. Ramadan 2003

The Master of Applied Science in Civil Engineering Program is a joint program with the University of
Ottawa, administered by the Ottawa-Carleton Institute for Civil Engineering

National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisitions et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-88894-0

Our file Notre référence

ISBN: 0-612-88894-0

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Canada

To my mother, the burning candle whom her light kept me from straying from my goals. She made everything I stand for.

To the soul of my late father, whom his sincere advice, understanding, guidance, and love are missed. Though he is not in this world, I always feel his presence. Whenever I gave up, I heard his encouraging and enthusiastic voice, calling for me to go on.

To my lovely fiancée Rania, whom her love and encouragement gave me hope for a better tomorrow.

Abstract

Transportation is a basic daily activity for almost all members of the society. However, it has been one of the main reasons for the loss of tens of thousands of lives and millions of injuries each year. Traditional safety analysis is performed using statistical or analytical methods. Such tools and techniques are project specific and therefore cannot be used on a network level. The advantages of having one tool that can integrate the different road safety techniques and procedures on state or provincial network level are therefore obvious. A GIS-based safety information system, named *Carleton University Safety Information System (CUSIS)*, has been developed. It is intended to provide a comprehensive analysis of existing rural highway networks. In its current version, CUSIS consists of five modules, namely: *Collision Analysis Module (CAM)*, *Highway Characteristics Module (HCM)*, *Operational Characteristics Module (OCM)*, *Consistency Evaluation Module (CEM)*, and *Video View Module (VVM)*.

Acknowledgements

I would like to express my sincere gratitude to three of my colleagues, who for their support, time, and effort enabled me to test and validate the developed system. In particular, I would like to thank Mr. M. M. El Shafey, Mr. M. E. Sarhan, and Ms. S. J. Taylor.

I would like to thank Ms. Heather McAdam, whom for her time, support, and effort; I was able to get the necessary GIS data. In addition, MTO provided necessary data and highway information. For this, I would like to thank Mr. B. Boutilier and Mr. D. Edwards, who for their knowledge, time, and effort, I was able to understand and analyse the provided data.

I would like to express my sincere gratitude for Prof. A. O. Abd El Halim, whom for his advice, moral support, and that encouraging pat on the shoulder; I was able to make it through my studies. His teachings will remain engraved in my mind.

I would like to thank and express my sincere gratitude to the thesis supervisors, Prof. Y. Hassan and Prof. D. E. Patterson, who for their invaluable time, patience, tolerance, support, and understanding, I was able to deliver this work.

Data entry was facilitated by assistants hired through the Ontario Work-Study Plan (OWSP) at Carleton University.

Financial support by Carleton University and the Natural Sciences and Engineering Research Council (NSERC) is gratefully acknowledged.

Table of Contents

Acceptance Form	iii
Abstract.....	iv
Acknowledgements	v
Table of Contents.....	vi
List of Tables	xi
List of Illustrations	xii
List of Abbreviations and Acronyms	xv
Chapter 1: Introduction	1
1.1. Introduction	1
1.2. Objectives	2
1.3. Scope of Research.....	3
1.4. Thesis Organisation and Writing Style	4
Chapter 2: Literature Review	5
2.1. Traditional Safety Analysis Methods and Systems	5
2.1.1. Statistical and Analytical Techniques	6
2.1.1.a. Before and After Evaluations	6
2.1.1.b. Identification of Hazardous Locations.....	7
2.1.1.c. Cost-Benefit Analysis in Development of Countermeasures.....	9
2.1.1.d. Analysis of Collision Trends.....	10
2.1.1.e. Collision Rate Comparisons of Locations with Different Features....	10
2.1.1.f. Cross-Sectional Evaluations	12
2.1.1.g. Comparison Group Evaluations	15

2.1.1.h. Risk Estimation/Analyses/Evaluation.....	15
2.1.1.i. Other Statistical and Analytical Techniques.....	17
2.1.2. Computer-Based Methods and Systems.....	27
2.1.2.a. Artificial Neural Network Systems.....	27
2.1.2.b. Fuzzy Logic Systems.....	29
2.1.2.c. System Dynamics Models.....	30
2.1.2.d. Expert Systems.....	31
2.1.2.e. Safety Information Systems.....	33
2.1.2.f. Computer-Based Safety Design Tools.....	36
2.2. Geographic Information Systems.....	39
2.2.1. GIS Defined.....	40
2.2.2. The Potentials and Capabilities of GIS in Transportation.....	41
2.2.2.a. Display/Query Analysis.....	43
2.2.2.b. Spatial Analysis.....	44
2.2.2.c. Network Analysis.....	45
2.2.2.d. Grid-Based Analysis.....	46
2.2.3. Existing GIS-Based Safety Analysis Methods and Systems.....	47
2.2.3.a. Commercial and Retail Systems.....	47
2.2.3.b. Researcher-Developed Systems.....	49
2.3. Conclusion: The Need.....	57
Chapter 3: Framework Development.....	59
3.1. The Vision.....	59
3.2. Framework Development and Interface Design.....	59

3.2.1. Frameworks Classification	61
3.2.2. Framework Design	62
3.2.3. User Interface Design	63
3.3. System Analysis: Identifying User Requirements	65
3.3.1. Abstraction.....	66
3.3.2. Encapsulation	69
3.3.3. Inheritance	70
3.3.4. Association.....	70
3.3.5. Communication with Messages	71
3.3.6. Infusing Methods of Organisation.....	71
3.3.7. Scale	71
3.3.8. Categorising Behaviour	72
3.4. Framework Architecture.....	72
3.5. Framework Components and their Functions.....	73
3.5.1. Collision Analysis Module (CAM)	74
3.5.2. Highway Characteristics Module (HCM).....	75
3.5.3. Operational Characteristics Module (OCM).....	77
3.5.4. Consistency Evaluation Module (CEM).....	79
3.5.5. Video View Module (VVM).....	82
3.6. Choosing the GIS Environment	83
Chapter 4: Data Collection, Formatting, and Processing	85
4.1. Introduction	85
4.2. Highway Data Referencing Systems.....	85

4.3. Data Collection	89
4.4. Data Reformatting	95
4.5. Integrating Data into ArcView® GIS	101
4.6. Data Processing	102
Chapter 5: CUSIS Synthesis	107
5.1. Introduction	107
5.2. “Collision Database” Application.....	110
5.2.1. “Plot a New Collision Table” Function.....	110
5.2.2. “Query Collision Database” Function	114
5.2.3. “Add a New Collision Record” Function.....	117
5.2.4. “Delete a Collision Record” Function.....	119
5.3. “Collision Analysis” Application	119
5.4. “Alignment Data” Application	125
5.5. “Traffic Volume” Application.....	128
5.6. “Raw Video Clips” Application	129
5.7. Building a Relational Database	130
Chapter 6: CUSIS Implementation.....	132
6.1. Introduction	132
6.2. Collision Analysis	132
6.3. Querying the Databases.....	134
6.4. Viewing Highway Videos	139
6.5. Suggested Countermeasures	140
Chapter 7: Summary, Conclusions, and Recommendations.....	141

7.1. Summary	141
7.2. Conclusions	142
7.3. Recommendations for Future Work.....	144
References	147

List of Tables

Table 4.1: Collision-data tables field description.....	91
Table 4.2: Highway alignments received from MTO.....	93
Table 4.3: Description of videotaped horizontal curves.....	94
Table 4.4: A table showing field name change for collision data tables	98
Table 4.5: Traffic volume data-table structure.....	99
Table 4.6: Table structure for elevation data.	100
Table 4.7: Table structure for horizontal curves data.....	100
Table 4.8: Attributes of “Data HWYs” theme.	104

List of Illustrations

Figure 3.1: An object featuring encapsulated variables by their methods (SUN 2003)....	70
Figure 3.2: CUSIS whole-part structure notation.....	73
Figure 3.3: CUSIS modular structure.....	73
Figure 3.4: Collision Analysis Module structure	75
Figure 3.5: Highway Characteristics Module structure.....	77
Figure 3.6: Operational Characteristics Module structure.....	78
Figure 3.7: Consistency Evaluation Module structure	82
Figure 3.8: Video View Module structure	82
Figure 4.1: Route Milepost System (Smith, Harkey, and Harris 2001, fig. 5).	86
Figure 4.2: Route-Reference Post system (Smith, Harkey, and Harris 2001, fig. 6).	87
Figure 4.3: Link-Node system (Smith, Harkey, and Harris 2001, fig. 7).....	88
Figure 4.4: A screen shot showing a view created and the Arc/Info coverages added to ArcView® GIS environment.	97
Figure 4.5: A screen-shot showing data layout and organisation after processing.	105
Figure 4.6: Portion of the digital map after processing. Scale 1:500000.	106
Figure 5.1: A screen shot showing the CUSIS menu within ArcView® GIS.	109
Figure 5.2: “Collision Database” main dialog.	110
Figure 5.3: Choice of highway to plot.	111
Figure 5.4: Collision distribution along highways 7, 15, 17, and 41 (MTO eastern region). Scale 1:1,370,038.....	113
Figure 5.5: Collision locations at intersection of highways 7 and 41. Scale 1:100,000..	114
Figure 5.6: “Collision Database -- Query” dialog.....	115

Figure 5.7: “Query results” dialog.....	116
Figure 5.8: Results export dialog.	117
Figure 5.9: “Collision Database -- Add” dialog.....	118
Figure 5.10: “Enter Vehicle Events” dialog.....	118
Figure 5.11: “Delete a collision record” dialog.....	119
Figure 5.12: “Password” prompt.	119
Figure 5.13: “Collision Analysis Method” dialog.....	120
Figure 5.14: “Choose Scope of Analysis” dialog.....	121
Figure 5.15: “OP Factors” dialog.	122
Figure 5.16: “Severity Weightings” dialog.....	124
Figure 5.17: Analysis Results dialog.	125
Figure 5.18: “Alignment Data - Choose Segment” dialog.	126
Figure 5.19: “Plot Options” dialog.....	127
Figure 5.20: “XY Graphing” dialog.	128
Figure 5.21: “Choose Location and Year” dialog.	129
Figure 5.22: “Video View Module -- Raw Videos” dialog.	130
Figure 5.23: Relationships between data tables.	131
Figure 6.1: Collision analysis results sorted by POI in a descending order.	133
Figure 6.2: Intersection of highways 41 and 132.	134
Figure 6.3: Collision database query results for Highway 41.....	135
Figure 6.4: Distribution of initial impact type.	136
Figure 6.5: Distribution of first event for single vehicle collisions.....	136

Figure 6.6: Road surface conditions for running off road, and skidding/sliding collisions.	137
Figure 6.7: Light conditions for collisions due to hitting a wild animal.	137
Figure 6.8: Horizontal and vertical alignments for LHRS 29650.	138
Figure 6.9: A photo captured from one of the videos, showing the mountainous terrain and the close trees.	139
Figure 6.10: A photo captured from one of the videos, showing one of the sharp horizontal curves.	140

List of Abbreviations and Acronyms

AADT	Average Annual Daily Traffic
ANN	Artificial Neural Network
CAM	Collision Analysis Module
CART	Classification and Regression Trees
CEM	Consistency Evaluation Module
C-LTPP	Canadian Long-Term Pavement Performance Program
CUSIS	Carleton University Safety Information System
DCRB	Digital Cartographic Reference Base of Ontario
EB	Empirical Bayes
FHWA	Federal Highway Administration (USA)
GIS	Geographic Information System(s)
GPS	Global Positioning System(s)
GUI	Graphical User Interface
HCM	Highway Characteristics Module
HSIS	Highway Safety Information System
km	Kilometres
LHRS	Linear Highway Referencing System
LRS	Linear Referencing System
MTO	Ontario Ministry of Transportation
NHS	National Highway System
OCM	Operational Characteristics Module
PDO, PDOnly	Property Damage Only

POI	Potential for Operational Improvement
qtd.	Quoted by/in
RLR	Red Light Running
RSA	Road Safety Audit
SPF	Safety Performance Function
UI	User Interface
UK	United Kingdom
US, USA	United States, United States of America
<i>v/c</i>	Volume to Capacity Ratio
VVM	Video View Module

Chapter 1:

Introduction

1.1. Introduction

Since the first evolvment of civilisation and humans are familiar with roads. However, it was not until the invention of the automobile, that roads took their current form. In the 1930's the earliest highway design guides and manuals appeared (Hassan 2002). Since then, highway design has been a growing concern for researchers, engineers, professionals, and highway officials whose intent is to improve, enhance, and sustain an acceptable set of guides. As the automotive industry evolved, faster cars appeared, collisions increased, and researchers have been keen on developing what is known as "highway safety explicit design."

It has been reported (NRC [US] 2001) that the number of people killed in traffic collisions in the last century in the United States of America is almost five times the number killed in all of the wars that this nation has been involved in since the year 1775. In the year 1998 alone, the number of people killed on United States highways was almost fifty-nine times the number of those killed in airplane crashes. In terms of frequency, a fatal highway collision occurs every thirteen minutes, and a highway injury occurs every fifteen seconds. Transport Canada (2001) indicated that more than 2,900 Canadians were killed and another 227,000 were injured in traffic collisions during the year 2000. On a local scale, nearly thirty percent of Ontario's traffic collisions lead to fatalities¹ or injuries (Transport Canada 2000), thus contributing heavily to the overall

¹ A fatal collision is defined as a collision in which at least one person sustains bodily injury resulting in death within 30 days of the collision (Transport Canada 2000).

preventable injuries. Almost 85,000 persons are injured or killed annually in traffic collisions. In the year 2001, 7,473 people were admitted to Ontario hospitals for 82,900 hospital-days due to traffic collisions; hence, preventing or reducing collisions on Ontario roads would have considerable social and economic impacts and could considerably relieve an overstressed healthcare system (Province of Ontario 2003).

Through the past years, significant progress has been made in the development and application of proper statistical methods in highway safety analysis to hold non-ideal conditions that often arise, and which cannot be handled by conventional statistical methods. Although the application of these newer methods is increasing, there is concern that the required tools and information to, adequately, undertake these analyses are not readily accessible to practicing highway safety professionals. On the other hand, widespread availability of statistical software has increased the risk of misapplying statistical techniques in safety investigations. (Persaud 2001)

Because of this struggle, evolves the practical need of a safety information system that integrates statistical techniques and procedures along with existing data records for collisions, traffic, and road characteristics. Such an information system would help highway professionals and officials direct their resources into the locations with the highest potential for safety improvements, and achieve the highest level of safety on roads under their jurisdiction.

1.2. Objectives

The objective of this thesis is to provide an innovative and integrated *highway safety information system*. The proposed system would integrate statistical and analytical

safety improvement techniques with data records for collision history, traffic volume, and highway characteristics. Within the proposed system, the intended user (i.e., transportation safety professional) can have an overall view of the state of practice for a certain highway or highway network. This system will address three key features missing in almost every safety improvement technique or procedure. The system must be:

1. generic, and highway independent;
2. having network analysis capabilities; and
3. able to establish a logical ranking of candidate segments for safety improvements depending on the total budget allocated.

The intended contributions of this thesis to the highway safety-engineering field are:

1. Unifying the diverse safety analysis tools into an integrated highway safety information system.
2. Applying data fusion on safety-related data to facilitate utilisation in safety analysis.
3. Eliminating the discrepancy between the state-of-the-art and the state-of-practice in highway safety engineering.

1.3. Scope of Research

The scope of this thesis is limited to two-lane two-way rural highways. The application of the proposed system was limited to four highways located in the eastern region of Ontario, namely: King's Highways number 7, 15, 17, and 41.

1.4. Thesis Organisation and Writing Style

This thesis consists of seven chapters. Introductory and background material is presented in this first chapter. *Chapter 2* provides a comprehensive literature review, where the author discusses the state of completed and on-going research of relevance. The intent is to provide an overview, while being as comprehensive as possible. *Chapter 3* focuses on the design of the framework used later to develop the intended safety information system. *Chapter 4* describes the procedure followed for data collection, reformatting, and processing. *Chapter 5* discusses the proposed system development. Within this chapter, the development environment is discussed, showing the justification for the environment choice and its potentials. *Chapter 6* discusses the system implementation, through diagnosing Highway 41. *Chapter 7* includes the summary, conclusions, and recommendations. A synopsis of reference material is provided.

Two matters of style should be mentioned. First, this thesis was written using Carleton University guidelines and the *Chicago Manual of Style*. The guidelines and the manual of style dictated margin size, line spacing, page numbering, preparing the list of references, parenthetical documentation, and abbreviations. In addition, for the sake of spelling consistency, *the Canadian Oxford Dictionary* was used as a reference for spelling. Second, the terms “collision”, “accident”, and “crash” are used synonymously. In addition, the terms “road” and “highway” are used synonymously. In addition, the acronym *HCM* refers to *Highway Characteristics Module* (discussed later), not to the *Highway Capacity Manual*.

Chapter 2:

Literature Review

A basic step in conducting research work is a comprehensive literature review, to ensure that this research is up to date, and to start from what others have reached. In conducting the literature review, there were four research areas to cover:

1. Current safety analysis or information systems.
2. *Geographic information systems* (GIS) technology and its application to safety information systems.
3. State of practice, and obstacles facing safety researchers to implement GIS technologies.

In addition, a study of GIS-applications framework development was conducted, and the findings are presented in the next chapter (*Chapter 3*). Past research-work has been covered and the findings were formatted in the following context.

2.1. Traditional Safety Analysis Methods and Systems

Literature is abundant on safety analysis methods and systems. At large, they can be classified as statistical, computer based, GIS based, and knowledge based systems. Due to the nature of this type of research, researchers are keen to invent, enhance, or modify safety analysis methods and/or systems. What follows is a comprehensive summary of the literature on each class of safety analysis methods and systems.

2.1.1. Statistical and Analytical Techniques

The literature is full of articles on collision analysis, traffic records, and statistical methods used to perform collision analysis and/or prediction. Persaud (2001) presented a comprehensive review of the current statistical and analytical techniques being implemented throughout various highway agencies. He conducted a survey where transportation jurisdictions were asked to report the statistical techniques in use. His survey resulted in eight groups of techniques and procedures (Persaud 2001, 4), namely:

1. Before and after evaluations.
2. Identification of hazardous locations.
3. Cost-benefit analysis in development of counter measures.
4. Analysis of collision trends.
5. Collision rates comparisons of locations with different features.
6. Cross-sectional evaluations.
7. Comparison group evaluations.
8. Risk estimation/analyses/evaluations.

Almost, every available statistical technique, or analytical technique, could be categorized into one of these categories. For those statistical or analytical methods that do not fit into one of these categories, a ninth category will be introduced in the following sub-sections.

2.1.1.a. Before and After Evaluations

These types of evaluations are conducted to assess the safety effectiveness of a given type of improvement or an improvement program as a whole. The information

obtained provides feedback to the process of planning future safety improvements. These studies range from simple before and after comparisons of accident counts to the more complicated Empirical Bayes (EB) approaches. For example, Griffith (2000) conducted a simple before and after study using comparison groups to study the safety effect of rumble strips on freeways. In addition, Hauer (1997) presented a detailed view of the whole process. Persaud et al. (2001) used the EB approach to evaluate the conversion of conventional intersections to roundabouts.

Al-Masaeid (1997, 369) compared before and after methods with EB methods. He concluded that before-and-after evaluation overestimates the safety benefits and results in erroneous conclusions at specific locations as well as at aggregate levels. Furthermore, he recommended that the EB method be used in evaluating safety improvements if there is a difficulty in identifying a suitable and large number of comparison locations.

2.1.1.b. Identification of Hazardous Locations

This is considered as the initial stage of the process through which locations are selected for safety improvement. Typically, the safety related data records of a location are used to identify and rank highway locations that should be investigated for safety deficiencies and possible treatment. The process is sometimes referred to as “Black Spot Identification.” More recently, the term “Identification of Sites with Promise” has been used (Hauer 1996c). Among those who used this unique term was Hauer (1996c, 54), where he reviewed historical and conceptual development of such procedures. Based on his review (1996c, 59), he attempted to create some order of thinking and made suggestions to improve identification of hazardous highway locations—that are:

1. joining all important clues not only collision history;
2. more attention is to be devoted to spotting collision clusters; and
3. using the entire collision history of a site not just a few years.

However, Hauer (1996c, 59) did not address the stage that follows identification—that is, the stage of diagnosis and remedy.

Various methods or techniques were used to identify hazardous locations. For example, Stokes and Mutabazi (1996) used the *Rate-Quality Control* method to identify hazardous locations. In addition, Sayed and Abdelwahab (1997) introduced a procedure to identify collision prone locations, or hazardous locations. They categorized collisions into one or a combination of the three highway system components and they used the collision membership to represent the correct ability of collisions by highway improvements. Furthermore, Persaud, Lyon, and Nguyen (1999) introduced an EB procedure for ranking sites for safety investigation by potential for safety improvement. A refinement of the EB method that is conceptually sound and inherently simple to use was the focus of their research. They provided a comparative evaluation of the proposed method against other EB methods. Their refined EB method was shown to be comparatively efficient. Moreover, Tarko, Sinha, and Farooq (1996) presented an area-wide safety analyses to detect areas that really need safety treatment. Their method used regression models to estimate the normal number of crashes in individual counties. Those numbers were used to predict the above-norm value for future years when a safety treatment is to be applied.

In addition, Davis and Yang (2001) combined Hierarchical Bayes methods with an induced exposure model in order to identify intersections where the collision risk for a given driver subgroup is relatively higher than that for some other group. Their computations were carried out using Gibbs¹ sampling, producing point and interval estimates of relative collision risk for the specified driver group at each site in a sample.

Persaud and Nguyen (1998) used an advanced method to identify unsafe intersections and evaluate the effect of a treatment within an Empirical Bayesian framework. Their models, unlike other models, allow safety estimates to be disaggregated by time, collision severity, impact type, and collision pattern. Finally, Persaud and Dzbik (1993) presented a distinctive approach, where generalised linear regression model estimates of a highway section collision potential and an EB procedure was used for refinement. The sections with the highest potential were the ones recommended for further investigation.

2.1.1.c. Cost-Benefit Analysis in Development of Countermeasures

This process involves the estimation and comparison of the costs and benefits of the available ways of remedying safety problems diagnosed at a specific location. This process not only ensures that only cost-effective measures are implemented, but it also facilitates the ranking of measures at a location and the ranking of all possible improvements in a jurisdiction, given the usual budgetary and other resource constraints (Persaud 2001, 4). Miller et al. (1998) considered four methods to allocate the costs of motor vehicle collisions between different vehicle types. They considered different vehicle types, corresponding to different perspectives, including that of the occupants of a

¹ Gibbs sampling, is a general method for probabilistic inference. (Lowry 2003)

vehicle and that of society under different property right assignments. However, there is an on-going dispute among researchers concerning the estimated value of a safety hazard or an improvement, and assigning to it a dollar value.

2.1.1.d. Analysis of Collision Trends

The analysis of collision trends has multiple objectives. This analysis could reveal patterns in collision experience that may indicate that specific highway features should undergo safety investigation, or that particular types of collisions should be targeted for countermeasure development. Also studying time trends could be used to detect deterioration in safety related to specific features and collision types or to detect patterns indicating the payoff of investments in collision reduction in general or certain types of collisions.

Dissanayake, Lu, and Chu (1999) described time series models developed from regression analysis. The models were used to forecast collision rates in association with some selected population subsets. They stated (1999, 44), “Such models are useful for assessing the safety performance of the subsets because the problems inherent in subsets are different from those of the average highway population.”

2.1.1.e. Collision Rate Comparisons of Locations with Different Features

These comparisons are frequently done with a view to characteristic differences in collision rates to differences in highway features. Safety effect estimates for various improvements are often obtained in this way (Persaud 2001, 5). A collision rate is used to normalise for differences in exposure to risk, e.g., in traffic volumes, between locations.

These studies are primarily done where a study of collision experience before and after an improvement is believed to be not viable.

Also, Hauer (2001, 69), addressed two questions concerning computing and interpreting collision rate. The first question that he addressed was, "Which of the many rates that have been proposed for measuring the safety of certain vehicle types of driver groups is to be trusted?" The second question was, "What meaning can be attributed to a finding of overrepresentation?" Furthermore, in the same article, Hauer discussed the correct and incorrect definitions of collision rate. He made a simple, yet an excellent, argument.

"It is concluded that apples and oranges should not be mixed. For a rate to be correct, the numerator and the denominator must pertain to the same entity." (Hauer 2001, 73)

Hauer also represented a unique answer to his second question concerning overrepresentation. He stated (2001, 73) that overrepresentation may be caused by a mix of three probability factors; "[...] the probability to be in an accident per unit of exposure, the probability of the accident to be reported, and the probability of an accident to be of the specified severity."

The use of collision rate as a measure has its deficiencies, for example, it produces biased results when applied to low volume roads (Hassan 2002). Thus, a decreasing collision rate may deceive the safety professional and gives him/her a false indication of safety improvement. Awakened by this fact, most researchers shifted to using collision frequency rather than rate. In another article, Hauer (1996b) developed a procedure to detect sites that are experiencing safety deterioration. He compared collision

frequency if it increased over time to more than what can be attributed to changes in traffic or to general trends. He used computer software to perform a statistical test of significance, to test for gradual and sudden increase in collision frequency. In conclusion, he emphasised that his procedure requires the exercise of judgment. He suggested a procedure to base such judgments on two practical considerations; the number of sites that can be practically subjected to further examination, and the yield margin of correctly identified sites.

In another article, Hauer (1996a) presented a primer on the testing of some common statistical hypotheses in highway safety. In this article, he applied the presented method to the specific case of testing statistical hypothesis about a change in the expected collision frequency other than that is the result of a change in traffic and similar influences.

2.1.1.f. Cross-Sectional Evaluations

Cross-sectional evaluations are undertaken to obtain safety effect estimates for various possible improvements; for example, there is the simple comparison of collision rates as summarised previously. Another example was presented by Sebastian (adapt. Persaud 2001), where he examined the collision rates of signalised intersections in Wisconsin with various types of left-turn treatments. He concluded that fully protected left turns, are the safest and that protected/permissive phasing is less safe than permissive only. Cross-sectional evaluations can also take the form of complex modeling in which collisions are first related in a regression equation to a variety of highway features, including traffic volume. The safety effect of making a change in one or more variable

can then be estimated using the equation to calculate the resulting change in collisions (Persaud 2001).

Council and Stewart (2000) used cross-sectional evaluations to evaluate the safety effects of converting rural two-lane highways to four lanes based on regression equations relating collisions to average annual daily traffic (*AADT*) for highways with these two types of cross sections. In addition, Mouskos et al. (1999) conducted a statistical analysis of the effect of various traffic, geometric, and environmental factors on collision rates on New Jersey State highways. They evaluated the effect of individual factors such as median, shoulder, number of lanes, and speed limits with the *Kruskal-Wallis*² test. They developed two different regression models that have sufficiently good (R^2)³ values. However, they recommended using these models with caution, without giving a clear justification for such a recommendation.

In another article by Zegeer et al. (1994), an analysis was performed to quantify the collision effects of shoulder widths on rural highways carrying fewer than two thousand vehicles per day. They used analysis of covariance to quantify collision relationships on low volume highways. They found that wider shoulders tend to lower collision rates. Another interesting finding by Zegeer et al. (1994) was that roads with a lane width of 3.0 meters and narrow shoulders had higher collision rates than roads with 2.7 meters lanes with narrow or wide shoulders. They also found that for traffic volumes greater than 250 vehicles per day, paved roads have significantly lower collision rates

² A non-parametric alternative to the one-way independent-samples ANOVA (Lowry 2003)

³ R^2 is the coefficient of determination. For further explanation, see Tabachnick and Fidell (2001).

than unpaved roads. Their research findings indicate that for low volume roads lane widths as narrow as 2.7 meters may be acceptable under certain conditions.

In a later study, Zegeer and Council (1995) summarised the known relationships between collision experience and cross-sectional highway elements. They also accounted for expected collision reductions due to related highway safety improvements. The elements they studied included lane width, shoulder width, shoulder type, roadside features, bridge width, median design, and other related elements. Furthermore, they concluded that widening lanes could reduce related collisions (i.e., run off road, head on, opposite direction sideswipe, and same direction sideswipe) by as much as forty percent. They also concluded that shoulder widening reduces related collisions by up to forty-nine percent. In addition, they concluded that bridge widening could reduce bridge-related collisions (i.e., collisions happening on a bridge) by up to eighty percent.

Again, the issue of shoulders and their effect on highway safety was raised by Polus, Livneh, and Katznelson (1999). They studied the flow and safety benefits of paved shoulders on heavily trafficked two lane rural highways. They developed a model to assess the impact of the use of shoulders by slow moving vehicles along the highway. In addition, they analysed speed data and collision rates for sections where the shoulders were recently paved. They found that speeds increased and collision rates were reduced at seventy percent of the sites studied. Finally, they concluded that the use of shoulders as a mean of increasing capacity and level of service is only valid if the shoulder width is at least 3.0 meters. They concluded by stating that using shoulder widening or paving to improve safety is only an interim solution.

2.1.1.g. Comparison Group Evaluations

These types of evaluations involve assessments of the suitability of untreated sites for use in a comparison group in before and after studies. The comparison group is used to control for other factors that may cause a change in safety performance when an action is employed. The intent is to separate the change in safety due to the treatment from the changes due to other factors. Griffith's study (2000) for the safety effect of rumble strips on freeways could fall under this category.

2.1.1.h. Risk Estimation/Analyses/Evaluation

This is a process for controlling (i.e., measuring, monitoring, comparing, and evaluating) levels of risk. It is done through an integrated series of steps that include combining collision data with exposure-to-risk data in order to calculate highway safety performance indicators, weighing up the accuracy associated with the estimated indicators, interpreting the various safety performance measures, calculating effectiveness of countermeasures, and defining and applying techniques for measuring safety and economic benefits. The risk estimation methods can be used to measure the highway-safety performance levels for any highway-user, vehicle, highway infrastructure, environment, or other characteristic. (Persaud 2001, 5)

For example, Brown and Baas (1997) studied the seasonal variation in collision frequencies and rates as a function of severity. They analysed the collision records of Quebec and compared it to US collision records. They found out that between thirty to fifty percent of the collisions in Quebec were associated with severe meteorological conditions; including rain, snow, hail, and icy conditions. They deduced that such harsh

conditions would contribute to dramatic safety performance deterioration on Quebec highways. They found that the lowest numbers and rates of fatal and serious injury collisions occurred during the winter season, however, “property damage only” (*PDO*) collisions are the most frequent during winter season. After doing further analysis of the collision records, they concluded that:

1. the effect of winter conditions on highway safety performance is complex;
2. there is a negative correlation between monthly rates of fatal and *PDO* collisions; and
3. collision severity should be explicitly used to enhance allocation of resources to achieve better protection of populations from highway collisions.

In addition, Council, Mohamedshah, and Stewart (1997) studied the effect of air bags on the severity of off-road fixed object collisions, as indicated by the impact of air bags on severity indices for fixed objects. In an earlier study, they concluded that air bags might have reduced severity indices by thirty-five to seventy-five percent (adapt. Council, Mohamedshah, and Stewart 1997). They extended their work by including data from two other states and from additional years. They admitted that there were inconsistencies among the data included in their analyses; however, they concluded that the presence of an air bag would decrease the proportion of serious or fatal driver injuries by ten to thirty percent for point objects, forty to fifty percent for guardrails, and ten to twenty percent for other barriers. Furthermore, they referred the insignificance of other objects tested, to the sample size available. Other than their findings, their definitions of objects tested and their methodology is somehow ambiguous.

2.1.1.i. Other Statistical and Analytical Techniques

Literature is also abundant on other statistical and analytical techniques that may not fit under the previous eight categories, or may fit under more than one category. For example, Gebers (1998) conducted a number of regression analyses of driving record variables measured over a six-year period. The techniques he presented were ordinary least squares, weighted least squares, Poisson, negative binomial, linear probability, and logistic regression models. His objective was to compare the results obtained from several different regression models under consideration for use, in the in-progress California driver record study. Discussing his results, Gebers stated

“[...] it was shown in all models that increased accident involvement was associated with increased prior citation and accident frequencies, possessing a commercial driver license, being young, being male, having a medical condition on record, and having a driver license restriction on record.” (Gebers 1998, 79)

Furthermore, he concluded that the results indicated that different regression techniques did not lead to any greater increase in individual collision prediction beyond that obtained through application of ordinary least squares regression. He also stated that the use of ordinary least squares regression appears to be safe, if the sample size is extremely large.

Khan, Shanmugam, and Hoeschen (1999) explored the relationship between fatal, injury, and PDO collision frequencies and traffic volume, segment length, and vehicle miles travelled. For the purpose of their study, they used a generalised linear regression-modeling framework. They evaluated Poisson, negative binomial, Gaussian, and lognormal distributions in terms of their ability to model collision data from an interstate

highway corridor in Colorado. They found that, for injury, PDO, and total collisions, the Poisson regression with log-transformed predictors, modeled the data significantly well. On the other hand, for fatal collisions, the lognormal regression performed better than other regression models examined. Models based on data disaggregated by median segment length, and AADT or vehicle miles travelled performed better than aggregated models. In conclusion, they found that Poisson regression models outperformed the other regression models examined.

In addition, Vogt and Bared (1998) developed negative binomial and extended negative binomial models to predict collisions on segment, three-legged intersections, and four-legged intersections of two-lane rural highways. The variables they used include traffic, horizontal and vertical alignments, lane and shoulder widths, roadside hazard rating, channelling, and number of driveways. They featured the models along with summary statistics, goodness of fit measures, and cross-validation between states. They concluded that segment collisions depend significantly on most of the highway variables collected, while intersection collisions depend primarily on traffic. They recommended development of adjustment factors for different regions and times, and further development of extended binomial regression models.

In addition, Lamm et al. (1995) presented a practical safety approach to highway geometric design. They based their procedure on statistical investigations of speeds and design parameters in Europe, the Middle East, and North America. The procedure they presented provided relationships between geometric design parameters, driving behaviour, and driving dynamics. Their objective was to determine sound highway

alignments and/or detect poor designs, and positively influence the collision situation. They concluded that the establishment of operating speed backgrounds, and relation design backgrounds is important when establishing highway geometric design guidelines. Furthermore, they emphasised that the relation backgrounds for different design levels should be developed as demonstrated in their article. Finally, they concluded by emphasising the practical value of their procedure.

Stewart (1996) demonstrated the application of *classification and regression trees* (CART), as a classifier for a discrete valued response variable or as a regression model for a continuous response variable. He indicated that the advantage of CART, is its ability to include a relatively large number of independent variables and to identify complex interactions among these variables. He presented a brief description of the CART procedure, and illustrated its application as a classifier and as a regression model to highway safety analyses. One of the powerful advantages of CART is that it is non-parametric and does not require data transformations. Thus, it is not necessary to specify a model form. Finally, Stewart presented three examples where CART was used. The first two, where it was used as a standalone model, CART provided the output of interest directly (i.e., average injury cost, likelihood of serious or fatal injury). In the third example, CART was used as a preliminary tool for another type of model.

As well Mohamedshah, Chen, and Council (2000) studied the association of selected factors with *red-light-running* (RLR) collisions. Their study was done on an urban scale; however, it is could be applied on a rural scale. Despite the hypothesis that the majority of RLR collisions result from inadvertent driver error or intentional

violation; they think that very little is known about the possible contribution of the geometric and traffic characteristics of intersections to RLR collisions. Motivated by this thought, they stated that the objective of this study was to examine selected geometric characteristics of signalised intersections and their impact on RLR collision rates and to establish a relationship between them. They used data from the *Highway Safety Information System (HSIS)* database to conduct two types of analyses. First was the limited contingency table analysis to examine the similarities and differences between RLR collisions and all collisions at urban signalized intersections. Second, regression type models were developed to examine the effects of intersection characteristics on RLR collision frequencies. Their results showed that average daily traffic, width of intersection, and traffic signal actuation were important non-driver factors for RLR collisions. In addition, they found out that these results differed slightly when the RLR vehicle was entering from the higher volume mainline vs. the lower volume cross street.

Norin and Isaksson-Hellman (1995, 45, 53-55) presented a method whereby safety potential of a safety design feature in a certain collision configuration could be predicted before the system is exposed to real traffic conditions. They combined data from collision tests or mathematical simulations with traffic collision data and considered collision severity and occupant size parameters. The purpose of their method was to create a means to predict how a new system, or a component for a vehicle, can influence the risk of injury. They concluded that collision severity, occupant size, and seating position influence the risk of injury. Furthermore, they stated that information on the distribution of these parameters increased the accuracy of prediction. They recommended that this method could be generalised for other collision types; however, they put a

condition requiring a relevant collision severity measure and a laboratory measurement value, which correlate well with the injuries sustained by the occupants.

Mountain, Fawaz, and Jarrett (1996, 695, 705) developed and validated a method for predicting expected collisions. They addressed the problem of predicting collisions on main roads with minor junctions where traffic counts on the minor approaches were not available. Their study was based on data for 3800 km of highway in the UK including more than 5000 minor junctions. The highways studied consisted of both single and dual carriage roads in urban and rural areas. They used generalised linear modeling to develop regression estimates of expected collisions for six highway categories. Furthermore, an EB procedure was used to improve these estimates by combining them with collision counts. Using EB method produced unbiased estimates of expected collisions for high-risk sites. Based on their results, they concluded that collisions on highways are not proportional to traffic exposure (i.e., traffic flow and link length), as is commonly assumed. For all six types studied, collision frequencies were shown to be a nonlinear function of annual traffic flows. Finally, they emphasised that the EB method is superior because it was the only method, which produced unbiased estimates of the treatment effect at high-risk sites. Despite the fact that EB estimates gave a small reduction in the mean squared errors relative to the predictive models.

Furthermore, Zhou and Sisiopiku (1997) examined the relationship between *volume-to-capacity ratios* (v/c) and collision rates. They selected a twenty-six km segment of Interstate I-94 in Detroit, Michigan as a study segment. They calculated v/c ratios from average hourly traffic volume counts. Collision rates were derived from

hourly-distributed number of collisions in the same study period. The correlation between v/c values and collision rates followed a general U-shaped pattern. The study results indicated that collision rates were highest in the very low hourly v/c range, decreased rapidly with increasing v/c ratio, and then gradually increased as the v/c ratio continued to increase. They concluded that traffic conflict was a major contributing factor to high collision rates observed in the high v/c range. On the other hand, they identified night conditions and driver inattention as explanatory factors for the occurrence of high collision rates in the low v/c ranges. However, they recommended that additional research involving more extensive database is necessary before their findings to be generalised.

Apart from all of the methods that have been developed, some articles provided precautions and criticism, or addressed problems with implementation of statistical methods. For example, Morrall and Talarico (1994, 145, 150-151) described a research project that was conducted to determine the amount of *side friction* demanded and provided for a range of roadway curvatures, vehicle speed and types, and pavement surface conditions. They developed a set of equations to predict the implied value of safe *side friction*, the safe speed of the curve, and the margin of safety provided by the safe speed. They concluded that the current, *at the time of research*, design standards are quite conservative and provide “a more-than-sufficient” margin of safety for vehicle drivers.

Mercier et al. (1997, 37, 46) assessed whether age, gender, or both influenced severity of injuries in head-on collisions on rural highways. Their initial hypothesis was that older drivers and passengers are more susceptible to injuries than younger ones, because of physiological changes, and other changes due to aging. The results of logistic

regression analysis indicated that four separate factors were strongly related to injury severity. Individual variables tested included age of driver or passenger, position of vehicle, and form of protection used. They used hierarchical and principal components logistic regression models to verify the effect of age related factors on injury severity, meanwhile amplifying findings of exploratory stepwise logistic analysis. They concluded that the use of lap and shoulder restraints, seat belts, was more beneficial for men than for women, while deployed airbags seemed to be more beneficial for women than men. They also concluded that the oldest passengers and drivers would experience greater injury severity from head-on collisions than would the rest of the population.

On the other hand, Chowdhury, Garber, and Li (2000) addressed the problem of multi-objectivity in highway safety design. They presented a multi-objective technique that kept the objectives in their respective units and provided a set of solutions, rather than one, which is most logical for a multi-objective problem. They faced a complicated problem, that is—highway safety objectives are often not commensurable. The methodology used was interactive multi-objective resource allocation. They determined the relationship between the decision variables with probability of crashes after implementing countermeasures, and crash severity levels after improvements. They concluded that the interactive multi-objective resource allocation method could enhance highway safety decision making. They also provided a list of data input for that methodology to work correctly.

Pietrucha, Pieples, and Garvey (2000, 12, 18-20) are an example of researchers who evaluated and criticised current safety programs. They presented a critical review of

the current, *at time of research, Pennsylvania Road Safety Audit Pilot Program*. They defined the *Road Safety Audit (RSA)* as a process whereby a team-of-experts attempts to identify features of the highway-operating environment that could be potentially hazardous, and then tries to eliminate or alter these features during the different phases of design before the system becomes operational. Their work describes the research conducted in cooperation with *Pennsylvania Department of Transportation*, and the *Pennsylvania Transportation Institute* conducted research on the application of a RSA process in two *Pennsylvania Department of Transportation* districts. After assessing the RSA process effectiveness, they recommended that *Pennsylvania Department of Transportation* expand the RSA process. Furthermore, they provided an itemised list of recommendations that included selection of teams, selection of projects, general audit procedure, conduct of field views, and development and communication of findings.

As for precautions and criticism to statistical methods, Washington (1999, 1, 3-6) provided a review of the mechanics of statistical hypothesis testing, and addressed five common misconceptions found in transportation research. He stated that most statistical methods use hypothesis testing. He gave several examples of the use of hypothesis testing in transportation research.

“Analysis of variance, regression, discrete choice models, contingency tables, and other analysis methods commonly used in transportation research share hypothesis testing as means of making inferences about population of interest.” (Washington 1999)

Furthermore, he stated that researchers often oversimplify the concept of hypothesis testing, thus they fall into one or all of the following five misconceptions:

1. α is the most important error rate.
2. Hypothesis test results are unconditional.
3. Hypothesis tests conclusions are correct.
4. Statistical significance (not effect size) determines practical significance.
5. Non-significant hypothesis test results are unimportant.

After going into every one of these five misconceptions, Washington concluded that many of the misconceptions or pitfalls could be remedied with careful application, and interpretation of his findings. Furthermore, he provided five recommendations, which, *in his opinion*, would remedy the five misconceptions reported. He concluded by stating that the Bayesian techniques are very powerful since they overcome the conditional nature of hypothesis tests, because Bayesian statistics offer a way to compute the probability of the null hypothesis being true, given the data, which is the main goal of researchers.

Saccomanno, Nassar, and Shortreed (1996, 14, 21-22) are another example of researchers who criticised existing statistical methods. They assessed the reliability of different statistical highway-collision injury-severity models. The criteria used were goodness of fit, robustness of risk factor coefficients, and intuitive capability and consistency of output. The results of their analyses indicated that model reliability was not sensitive to the number of injury classes specified in the model or to the level of model aggregation; however, there was no indication that a significant transfer of error took place from one severity level to another in a sequential model structure. Furthermore, their results suggested that reliability of statistical highway-collision

severity models depends primarily on the accuracy of information provided in the collision data.

Criticising the current statistical methods applied by researchers; Miaou, Lu, and Lum (1996, 6, 13) demonstrated the pitfalls of using R^2 to evaluate goodness of fit of collision prediction models. They justified their argument by using computer simulations of commonly used collision prediction models. They emphasised their argument by stressing on non-normal and nonlinear nature of collision prediction models. Throughout their analyses, they provided three properties for an alternative measure for goodness of fit. Finally, they briefly reported two recent research efforts aimed at developing alternative measures with such properties, namely; the scaled deviance, and the *Akaike information criterion*⁴.

Finally, Mensah and Hauer (1998) discussed two problems of averaging arising from the use of a single *safety performance function (SPF)*, instead of using one for the daytime and another for nighttime. They defined SPF as a function that links the expected collision frequency to traffic flow. Their point of view of an ideal SPF was a function that represented cause-effect regularities. Their first argument was that collision counts were performed over a long period and because average flows were used, a problem of argument averaging exists. Their second argument was concerned with the problems arising if one joint SPF was estimated when two or more separate functions should have been used, which they defined as a function averaging problem. After discussing the

⁴ “Akaike's Information Criterion is a criterion for selecting among nested econometric models.” (Moffatt 2003)

consequences of argument averaging and the bias due to argument averaging, their only recommendation was to use a daytime and a nighttime SPFs separately.

2.1.2. Computer-Based Methods and Systems

Ever since computers were introduced to the public, transportation researchers were among the pioneers to use it. The accessibility of computers and their capability to perform several million instructions per second, gave a push towards exploring new frontiers in transportation research. Literature is not that prevalent as it would be on statistical methods. However, the approaches of transportation researchers varied from using artificial neural networks, system dynamics, fuzzy logic systems, and expert systems. In addition, there was great efforts towards building collision analysis and/or information systems. A brief review of such systems is illustrated.

2.1.2.a. Artificial Neural Network Systems

Artificial neural networks (ANNs) gained wide respect among researchers for their predictive ability. ANNs provide a means to mimic the human brain's computing architecture. They are analogous to a young child who learns as time goes by; in other words, trained through the arguments that he/she deals with. ANNs represents a distinct methodology that deals with uncertainty. ANNs can represent complex nonlinear relationships, and they are very good at classification of phenomena into pre-selected categories used in the training process. On the other hand, the outputs precision is sometimes limited because the variables are effectively treated as analog variables, and *minimization of least squares errors* does not necessarily mean *zero error*. Perhaps the weak point of ANNs is their need for generous data that are representative and cover the

entire range over which the different variables are expected to change. (Tsoukalas and Uhrig 1997, 2-5)

Harlow and Wang (2001) considered the development of a system for automatically detecting and reporting traffic collisions at intersections. The complete system would automatically detect and record traffic conditions associated with collision such as time of collision, video of collision, and the traffic light signal control parameters where applicable; however, they conducted basic research required to develop the system rather than developing the entire system. This involved developing methods for processing acoustic signals and recognising collision events from the background traffic events. They created a database of vehicle crash sounds, car braking sounds, construction sounds, and traffic sounds. The *Mel-frequency cepstral coefficients*⁵ were computed as a feature vector for input to the sound classification system. Finally, an ANN was used to classify these features into categories of crash and non-crash events. Their classification testing resulted in ninety-nine percent accuracy.

On the other hand, Zhang and Yang (1997, 287, 293) developed and applied a machine learning method—that is, *instance-based learning*, to highway collisions frequency predictions. The data set used contained collision data from the main Utah highways for a five-year period, 1988 to 1992. Their experimental results showed that the instance-based learning method is applicable to highway collision predictions and compared favourably with linear regression analysis. They concluded that instance-based learning methods could be one of the promising methods for this application.

⁵ *Mel-frequency cepstral coefficients* are used to parameterise speech and different voice patterns in speech recognition algorithms (Harlow and Wang 2001).

Furthermore, they expressed their commitment to applying this method to a larger highway collision dataset, and that they will conduct detailed analysis of the performance.

2.1.2.b. Fuzzy Logic Systems

Fuzzy logic systems address the imprecision of the input and output variables directly by defining them with fuzzy numbers, and fuzzy sets, that can be articulated in linguistic terms (e.g., cold, hot, happy, angry, dangerous, and so forth). Furthermore, they allow far greater elasticity in formulating system descriptions at the proper level of detail. This means that complex process behaviour can be described in general terms without precisely defining the complex, usually nonlinear, phenomena involved. Rephrasing *Occam's razor*⁶, we may say that fuzzy descriptions are more prudent and hence easier to formulate and modify, more biddable, and perhaps more lenient of change and even failure. (Tsoukalas and Uhrig 1997, 3)

Sayed, Abdelwahab, and Navin (1995, 352, 357) used fuzzy logic pattern recognition to identify collision prone locations. They realised that current methods to identify collision prone locations made no distinction between collisions that resulted from highway and non-highway related factors. They argued that by combining collisions that are treatable and non-treatable by highway improvements, could be misleading and could lead to misallocation of funds by highway authorities. They used fuzzy logic pattern recognition to categorise collisions into a finite set of categories, which are based

⁶ "Occam's razor is a logical principle attributed to the mediaeval philosopher William of Occam (or Ockham). The principle states that one should not make more assumptions than the minimum needed. This principle is often called the principle of parsimony. It underlies all scientific modelling and theory building." (Heylighen 1997)

on safety experts' knowledge. The proposed categories could include one or more of the three components of the highway system—that are, the driver, the vehicle, and the highway. They tested their method using data from the collision database of the *British Columbia Ministry of Transportation and Highways*. One can summarise their work as a method to identify collision prone locations by assessing collision-contributing factors. They achieved this by building a fuzzy logic system, which would assess the degree to which each collision belongs to the three highway system components. The output of their system was the degree of membership of each collision in the three highway system components. Finally, they interpreted their results as a successful indicator that this approach can effectively classify collisions into a finite set of categories.

2.1.2.c. System Dynamics Models

System dynamics has been developing for the last thirty-five years and now has a worldwide and growing acceptance. System dynamics combines the theory, methods, and philosophy needed to analyse the behaviour of systems not only in management, but also in environmental change, politics, economic behaviour, medicine, engineering, and other fields. System dynamics provides a common foundation that can be applied wherever we want to understand and influence how things change through time. The system dynamics process starts from a problem to be solved, a situation that needs to be better understood, or an undesirable behaviour that is to be corrected or avoided. The first step is to spout the wealth of information that people possess in their heads. The mental database is a rich source of information about the parts of a system, about the information available at different points in a system, and about the policies being followed in decision-making. System dynamics uses concepts adopted from the field of feedback control to organise

available information into computer simulation models. A computer as a simulator acts the roles of the operating people in the real system; hence, revealing the behavioural inferences of the system that has been described in the model. (Forrester 1991)

Mehmood, Saccomanno, and Hellinga (2001, 37, 46) presented a framework of a systems dynamics model for simulating highway collisions. They applied the model to two-vehicle rear end collision where both vehicles were assumed to be traveling in the same lane. The collision scenario involved the introduction of some obstruction along the pathway, which caused the drivers to adjust their speeds and separation distances. The system dynamics model extended the features of classical car following theory in conjunction with collision avoidance models. The system provided numerous insights into the complicated process that gave rise to rear end collisions for the assumed transportation scenario. Such insights were useful and could serve as a guide to developing effective safety countermeasures for reducing rear end collisions. Finally, they concluded that further calibration and validation of the model components were required; however, it was shown that system dynamics models were useful for evaluating changes in collision potential from the introduction of alternate safety countermeasures.

2.1.2.d. Expert Systems

Expert systems are computer programs that contain knowledge in a specific domain. The expert system often uses this knowledge to perform tasks that humans can do. There are different architectures for expert systems, for example, rule-based systems, and knowledge-based systems.

For example, Thielman and Griffith (1999) provided an overview of three expert systems for collision data collection. These systems were initially developed and being run by the *FHWA Crash Data Collection Expert System Program*. The three expert systems were implemented by using the rule-based expert system technology. The knowledge of experts was encoded into rules that were used to determine which data the system collects and to reach a conclusion that an expert would reach. The goal of the *FHWA Crash Data Collection Expert System Program* was to use expert systems technology to improve the accuracy and consistency of police reported data. In this program, police officers used pen-based computers that contain the expert systems to collect on-scene collision data. The experts systems used loaded knowledge to, intelligently, select data to collect. The expert systems were evaluated during two field tests. The results indicated that the expert systems were well accepted by the police officers. The officers collected data at an average of two minutes per expert system. Finally, Thielman and Griffith (1999) recommended using the expert systems to collect data for highway safety analysis, extending the scope of these expert systems, and developing additional expert systems to collect data for analysis.

On the other hand, Sayed (1997, 251, 265-266) presented a thorough description of the diagnosis phase of a highway safety expert system. The overall objective of the expert system was to provide highway safety officials with an efficient tool to identify collision prone locations. Then, quickly and reliably provide advice on the appropriate remedies based on the analyses of the collision and highway environment data. The proposed system had three basic phases—that were detection, diagnosis, and remedy. Within the diagnosis phase, a knowledge-based system was developed to identify causes

and the contributing factors of safety problems at collision prone locations, and to provide the appropriate remedies. He showed that the knowledge-based approach best suits the diagnosis phase as it involves great deal of judgement and experience by the safety professional. He addressed the issues of knowledge acquisition, problem solving strategy, system features, uncertainty handling, and system verification and validation. Finally, he concluded that the use of the knowledge-based approach within the diagnosis phase gave satisfactory results.

Unlike researchers who used rule-based or knowledge-based expert systems, Zhang, Back, and Zhou (2001, 476, 481) developed a safety expert system under the name *Traffic Accident Analysis System (TAAS)*, which used traffic rules (i.e., law) as knowledge source to judge if a driver was responsible for a traffic collision. TAAS separated knowledge base and inference engine, using production rule and backward chaining. In addition, TAAS used a predefined text and tracing program to realise the explanation mechanism. However, TAAS had a disadvantage—that was, data were stored orderly, such that the inference engine had to start from the beginning to search for a rule. This searching mechanism took, relatively, a lot of time in a way that undermined the performance of the system.

2.1.2.e. Safety Information Systems

Since computers can store and handle enormous amounts of data, and for their fast processing speed, they are the corner stone of most of safety information systems. Within a typical information system, all relevant data are stored and managed through a database management system. Various efforts have been made to go that extra mile, and

to process the data to get useful information, which could help safety professionals achieve a safer design for safer highways.

One of the most famous safety information systems is the one developed by the FHWA. The FHWA has developed a highway safety database that can meet the safety professionals' needs which is the *Highway Safety Information System (HSIS)*. The HSIS uses existing data that is already being collected by States (i.e., California, Illinois, Maine, Michigan, Minnesota, North Carolina, Ohio, Utah, and Washington) for the management of the highway transportation system, for the study of highway safety. The HSIS is a highway-based system that provides quality data on a large number of collisions, highway, and traffic characteristics. The data are acquired annually from select States, processed into a common computer format, documented, and prepared for further analysis.

Paniati and Council (1997) described and discussed the HSIS. They gave a detailed description of its data files, which are Accident, Roadway Inventory, Traffic Volume, Roadway Geometrics, Intersection, Interchange, Vehicle Identification Number, and Guardrail/Barrier data files. Furthermore, they discussed the HSIS evolution from the late 1980s until the date of publication in 1997. To transform data into knowledge, however, requires access to individuals with expertise in the analysis of highway safety problems.

Zegeer et al. (1998a, 1, 9) investigated the *National Highway System (NHS)* roads in the HSIS States. The purpose of their study was to investigate how HSIS might be used to examine NHS safety issues. The overall observation was that collision rates on

NHS roads were ten percent lower than non-NHS roads. Furthermore, for rural roads, fixed object collision rates were higher on NHS roads than on non-NHS roads. After analysing the HSIS data for NHS roads and after observing what was mentioned earlier, they concluded that highway designers and safety officials could use this type of information about collision rates and roadway characteristics to enhance safety by upgrading existing highways and improving design of NHS highways to some specified highway design guidelines.

Furthermore, the USA federal government maintains two databases for collisions: the *Fatal Accident Reporting System* and the *General Estimates System* (Souleyrette et al. 1998, 2). In addition, most of the states maintain a database for traffic collisions, e.g., Iowa's *Accident Location and Analysis system (ALAS)*, and Alabama's *Comprehensive Accident Rapid Evaluation system (CARE)* (Souleyrette et al. 1998, 5). Iowa's ALAS has been subject to serious modifications and upgrades, yet the system core remained almost the same and the modifications were all superficial. Furthermore, one can generalise the judgement that these systems have one or more of the following disadvantages:

1. They do not offer safety analysis features.
2. Output is not user friendly.
3. Running the system is cumbersome or tedious.
4. The system is not transportable and sometimes machine dependant.

On a Canadian scale, Leore (2000; 2003) has been working on *Canadian Highway Information System (CHIS)*. He used TransCAD® to develop the first federal highway information system in Canada. He has been developing this system since 1995.

His goal was to develop a comprehensive database of highway-related information for all parts of Canada through a cooperative data sharing arrangement with Provincial/Territorial Highways Departments. Although CHIS was designed to provide variety of data to transport research in general, Leore limited the scope of his application to policy analysis. However, CHIS is currently limited to displaying data and highway inventory, without providing any further analyses to collision data.

On a local scale, the *Ontario Ministry of Transportation (MTO)* maintains a number of databases for safety related data (Province of Ontario n.d.). These databases are, *Accident Information System*, the *Highway Inventory Management System*, the *Integrated Highway Information System*, and the *Traffic Volume Information System*, however, none of the acquired literature items had any explanation or description of these systems.

2.1.2.f. Computer-Based Safety Design Tools

The only tool mentioned in literature that can fit in this category is the FHWA *Interactive Highway Safety Design Model (IHSDM)*. The IHSDM is a suite of software analysis tools for evaluating safety and operational effects of geometric design decisions on two-lane rural highways. IHSDM provides highway project planners, designers, and reviewers in highway authorities and engineering consulting firms with a suite of safety evaluation tools to support these assessments. (Krammes and Hayden 2003)

Earlier this year, the FHWA released the 2003 release of IHSDM, which concluded a multiyear research and development effort. Highway-project decision makers now can use IHSDM to check designs for conformance with design policy,

estimate their expected safety performance, and diagnose potential safety and operational issues throughout the highway design process. IHSDM consists of six modules, which are crash prediction module, design consistency module, driver/vehicle module, intersection review module, policy review module, and traffic analysis module. (Krammes and Hayden 2003; USA 2003a)

The *crash prediction module* estimates collision potential for a design alternative, including all entered highway segments and intersections. Estimates are quantitative and include the number of collisions for a given highway segment or intersection as well as percentage of fatal, injury, and PDO collisions. The *design consistency module* evaluates the operating-speed consistency of two-lane rural highways and flags horizontal alignment elements that are inconsistent. The employed speed-profile model estimates 85th percentile speeds (V_{85}) on each element along an alignment. Although consideration of vertical alignment when predicting V_{85} was not clear in the applied method, the FHWA claims that vertical grades and curvature are being considered within their model.

Furthermore, the FHWA stated:

“The speed-prediction models allow highway designers to estimate the impact of horizontal curve radius and vertical grades/curvature on free-flowing passenger vehicle speeds along a two-lane rural highway” (USA 2003a).

The *driver/vehicle module* consists of a driver performance model linked to a vehicle dynamics model. Driver performance is influenced by the roadway/vehicle system, i.e., drivers modify their behaviour based on vehicle and roadway conditions. On the other hand, vehicle performance is affected by driver's behaviour and performance. The driver performance model estimates the driver's speed and path along a two-lane

rural highway in the absence of relevant traffic data. The resulting estimates serve as input to the vehicle dynamics model, which estimates corresponding measures including lateral acceleration, friction demand, and rolling moment. (USA 2003a)

The *intersection review module* evaluates the geometric design of at-grade intersections on rural two-lane highways and identifies safety problems. This module consists of a policy review component and a diagnostic review component. The policy review component checks the intersection design elements—that are, corner radius, turn lane design, intersection angle, and intersection sight distance triangles. The diagnostic review component is an expert system that leads the user through a systematic evaluation of an existing or proposed intersection geometric design to identify safety problems. The *policy review module* is designed to be used throughout all of the stages of highway planning and design, including design review for new or reconstruction projects, where design elements that do not abide by the policy are identified. The output includes a summary of the policy review, including a listing of all design elements that deviate from policy. The categories of design elements to be verified include horizontal alignment, vertical alignment, cross section, and sight distance. (USA 2003a)

Finally yet importantly, the *traffic analysis module* that links highway geometry data with a traffic simulation model to provide information on speed, travel time, delay and percent following in sections. *TWOPAS* (a traffic simulation model for two-lane rural highways) forms the foundation for this module. The TWOPAS model simulates traffic operations on two-lane rural highways by reviewing the location, speed, and acceleration of individual vehicles on the simulated road every second and advancing those vehicles

along the road in a pragmatic manner. The model takes into consideration the effects of road geometry, driver characteristics and preferences, vehicle size and performance characteristics, and the presence of opposing and same-direction vehicles that are in driver's sight at any given instant. (USA 2003a)

Despite of all these timely modules, IHSDM is still in the beta-testing phase and it still suffers from some major disadvantages, for example:

1. the user interface is not friendly;
2. the input mechanism is cumbersome, and the slightest mistake would result in a disfunctional data file;
3. IHSDM does not realise the segment-to-segment interaction, or the road-network interaction; and
4. IHSDM does not offer a wide variety of data exchange formats.

2.2. Geographic Information Systems

A transformation is taking place. Businesses and government, schools and hospitals, non-profit organisations, and others are taking advantage of it. All around the world, people are working more efficiently because of it. Information that was limited to spreadsheets and databases is being unleashed in a new way—all using GIS. Linking location to information is a process that applies to many aspects of decision making in business and the community. Choosing a site, targeting a market segment, planning a distribution network, zoning a neighbourhood, allocating resources, and responding to emergencies—all these problems involve questions of geography. (ESRI 2002)

2.2.1. GIS Defined

According to Patterson (2002, 1), GIS is a specialised computer software, which shares characteristics with spreadsheets, database, and computer-aided design programs. It is possible to analyse spatial information within these programs, however, you generally cannot observe the spatial pattern that may be evident. A GIS is capable of data input, manipulation, handling, analysis, and output. Historically, GIS software was categorised as either raster or vector-based. *ArcView® GIS* allows you to work with both categories of data. *ArcView® GIS* has a short learning curve and is extremely powerful from an analytical perspective. In addition, Harkey and Griffith (1999) defined GIS as a collection of hardware and software that is used to edit, analyse, and display geographical information stored in a spatial database.

In the past few years, many transportation jurisdictions and other related entities, have examined the feasibility of using GIS for transportation planning, management, and engineering applications. In some jurisdictions, GIS is being used to plan transportation routes, manage pavement and bridge maintenance, and perform various traditional transportation-related functions. In the transportation industry, geographic analysis is the key to making better decisions. Smith, Harkey, and Harris (2001, 3) summarised the common uses of GIS in transportation in the following points:

1. Pavement and bridge maintenance management;
2. modeling disaster response plans;
3. quantifying the potential impacts of transportation alternatives;
4. routing of overweight and oversized vehicles;

5. flood prediction;
6. risk assessment and risk management;
7. seismic slope-performance analysis and mapping of landslide hazard zones;
8. study of air emissions on health; and
9. truck traffic analysis for the management of rural highway networks.

By now, it is obvious that GIS serves three distinct transportation needs:

infrastructure management, fleet and logistics management, and transit management.

Transportation professionals use GIS to integrate mapping analysis into decision support for network planning and analysis, tracking and routing, inventory tracking, route planning and analysis. One area where GIS has not yet been extensively applied is highway safety-related applications. (Smith, Harkey, and Harris 2001; Souleyrette and Gieseeman 1999)

2.2.2. The Potentials and Capabilities of GIS in Transportation

There is no apparent reason for the limited, maybe perishing, use of GIS in highway safety analysis. In part, this may be due to a lack of understanding of the potential benefits of such an application. Thus, prior to developing a GIS highway-safety information system, there is a need to have a better understanding of how GIS can benefit traditional techniques and methods. The benefits of GIS are well recognised in a number of disciplines. GIS provides the capability of storing and maintaining large data sets of spatial and tabular information. GIS has its might in providing display and analytical capabilities that model the physical propinquity of spatial features. One powerful aspect of GIS is the flexibility in modeling spatial objects to suit the particular needs of the user

or application. These capabilities have evolved as the technology has matured. (Lang 1999; Lo and Yeung 2002; Smith, Harkey, and Harris 2001; Souleyrette et al. 1998)

In its early days, GIS provided elementary analysis capabilities for areas that were represented as discrete points distributed throughout a uniform grid. This type of analysis is referred to as grid analysis. GIS has since matured to include systems based on cartographic representation of points, lines, and area feature types. These systems provide a topological data model that allows for more analysis capabilities that is robust, referred to as vector-based analysis. Other common GIS capabilities include database integration; image overlay capabilities, and network analyses. Over the past few years, GIS has been bespoken to accommodate linearly referenced data. Collision and roadway inventory data are examples of this type of linear data and can now be brought into GIS for display and analysis. This capability offers the safety professional specific analytical methods for understanding the spatial relationship of data that are not found in other information systems. (Lang 1999; Lo and Yeung 2002; Smith, Harkey, and Harris 2001; Souleyrette et al. 1998)

In addition, GIS offers a programming or scripting environment that allows the user to develop specific analysis programs or customise existing programs. All functions for display and analysis can be employed in a single-system design for *Rapid Application Development (RAD)* using common programming languages, such as *Visual Basic*, *C++*, *Visual C++*, *C#*, and *Java*. This capability is palpable in the FHWA *GIS Safety Analysis Tools*, which were developed in *ArcView® GIS* using the *Avenue*⁷ programming language. It is even more evident in the safety information system developed and

⁷ ArcView object-oriented programming language. Developed by ESRI, Inc.

described in this thesis. More importantly, with recent developments in interoperability, GIS can be integrated into more mainstream enterprise applications, as well as web-based applications. (ESRI 1996a; Lang 1999; Lo and Yeung 2002; Smith, Harkey, and Harris 2001, 4; Souleyrette et al. 1998)

Furthermore, Smith, Harkey, and Harris (2001) praised GIS capabilities for providing the ability to display and view collision and highway inventory location, and offers great rewards not available in a *linear referencing system (LRS)* alone. According to them, this capability is broader than simply mapping data and includes several types of analytical capabilities that they broadly categorised into four groups. Namely:

1. Display/Query analysis.
2. Spatial analysis.
3. Network analysis.
4. Grid-based Analysis.

2.2.2.a. Display/Query Analysis

The primary appeal of GIS to users is the graphical capabilities. As it has been stated, “A picture is worth a thousand words.” Maps are the pictures GIS uses to communicate complex spatial relationships that the human eyes and mind are capable of understanding. The computer makes this possible, but user determines what data and spatial relationships will be analysed and portrayed, or how the data will be thematically presented to its intended audience. (Lo and Yeung 2002; Smith, Harkey, and Harris 2001, 4)

For linearly referenced data to be displayed in GIS environment, it first must be integrated with spatial data. GIS can integrate spatial data of various scales, resolutions, and projections, although use of spatial data integration warrants caution on inappropriate use. Hence, the user has to put in perspective the relative scale of data. Data integration provides a microscopic level of analysis through the ability to spatially integrate and merge the data into a single view. Data not ordinarily used by the safety engineer, data that would otherwise be external to the LRS or not have a linear reference, such as demographic data, meteorological data, environmental data, economic data, and terrain data, to name a few, can be integrated as tabular data within the GIS environment. LRS data that is not ordinarily integrated, such as work-zone data, can also be integrated in the same manner within GIS, therefore expanding the data sources available to the safety professional. (Lo and Yeung 2002; Smith, Harkey, and Harris 2001, 5)

2.2.2.b. Spatial Analysis

Overlay analyses are the techniques used within GIS environment for spatial analyses and data integration. GIS is resourceful of tools to combine data, identify overlaps across data, and join the attributes of data sets together using feature location and extent criteria. Overlay techniques would combine spatial data in other ways—that is, features can be combined to, simply, add one spatial data set to another, to update, or replace portions of one data set with another data set. Overlay analysis can be used to merge spatial data by combining two or more spatial data sets to produce a new spatial data set where the feature attributes are a union of the input sets. (Lo and Yeung 2002; Smith, Harkey, and Harris 2001, 5)

Proximity analysis is another type of GIS query capability and a category of spatial analysis that distinguishes GIS from all other information systems. Buffering is a means of performing this practical spatial query to determine the proximity of neighbouring features. In GIS, buffering will locate all features within a prescribed distance from a point, line, or polygon. Such as determining the number of crashes that occurred within a certain distance of an interchange, or locating secondary crashes that occurred within a certain distance and time of other crash events. However, reliability of these variables may not always support this example. (Lo and Yeung 2002; Smith, Harkey, and Harris 2001, 6)

2.2.2.c. Network Analysis

Network analysis is restricted to searching along a line, such as a route, or throughout a network of linear features, such as the road network. Network analysis can be used to define or identify route corridors and determine travel paths, travel distances, and response times. For example, network analysis may be used to assess the traffic volume impact of a road closure on adjacent roadways. To improve the network model and provide the capability of automated route selection, the road network can be developed to include turning points, avoid improper turns onto one-way streets, represent posted traffic control restrictions, and include impedance factors to travel (such as mean travel speeds, number of travel lanes, and traffic volumes) to enhance the network analysis. (Lo and Yeung 2002; Smith, Harkey, and Harris 2001, 7-8)

2.2.2.d. Grid-Based Analysis

Grid-based analysis uses a grid or cells to aggregate spatial data for discrete distribution. In grid-based analysis, the spatial data are developed as tiles of a given dimension, or points of a uniform distribution, as defined by the user, for display and analysis. Grid-based analysis is effective in displaying patterns over larger areas, such as representing the sum total of collisions that are located within a cell. This capability provides a quick means to view spatial clustering of collision data. This technique is favoured among highway authorities that assign collision data to highway midpoints and intersections, a method that, by nature, forms data clusters. Since grid-based analysis aggregates data at a specified grid resolution, it would not be appropriate for site-specific spatial analysis. In other words, it would be more efficient if used on a network-wide scale. (Lo and Yeung 2002; Smith, Harkey, and Harris 8-9)

In grid-based analysis, GIS provides a set of special tools that are available to merge grid data for overlay analysis. Grid-based overlay analysis is similar to the GIS overlay analysis discussed earlier; however, the techniques and functions available in grid-based analysis are to some extent different. When the cells of different data sets have been developed using the same spatial dimensions, they can be merged on a cell-by-cell basis to produce a resulting data set. The functions and processes used in grid-based analysis to merge grid data are referred to as *map algebra*, because the grid data sets in grid-based analysis are merged using arithmetic and Boolean operators called *spatial operators*. (Lo and Yeung 2002; Smith, Harkey, and Harris 8-10)

2.2.3. Existing GIS-Based Safety Analysis Methods and Systems

It was very difficult locating literature that discusses or describes GIS-based safety analysis tools; however, few were located, and represented within the context of the following paragraphs. To make it easier for the reader, and to improve the capability of fully understanding each method or system architecture, the existing systems were classified into two major categories, which are:

1. Commercial and retail systems.
2. Researcher-developed systems.

2.2.3.a. Commercial and Retail Systems

JMW Engineering, Inc. (2003) developed a GIS-based system to manage collision databases. They commercially distribute this system under the name *Accident Information Management System (AIMS)*. They claim that AIMS can plot collisions on a map in *three-dimensional* (3D) form, however, this is a misleading claim, as the three dimensions are not for collision location, but rather for graphical representation of the number of collisions; in other words, it provides 3D plots or pin maps. Another misleading claim is AIMS data cross-compatibility. The developer claims that AIMS could adapt itself to any form of data. In fact, AIMS could be customized, at the developer's end, to adapt the client's database. In summary, other than creating descriptive reports, plotting graphs, and querying database AIMS cannot offer any type of engineering-based safety analysis.

Crossroads Software (2003) is another commercial developer for *collision database management systems*. The commercial name of their product is *Traffic Collision*

Database (TCD). The TCD provides data input and management for collisions, citations, queries and reports, including historical, high incidence, and monthly, as well as collision reports by day and hour and other parameters; graphs and charts for such categories as highest degree of injury, collision type, weather and lighting conditions, and so forth. The developer claims that TCD is capable of analysing the data; however, not a single description of its analytical capabilities was mentioned in the information sheet provided by the developer. It is assumed that TCD analytical capabilities will be oriented for usage in traffic planning but will not offer any decision support for designers or policy makers.

In addition, there is a system developed by CalGIS under the name *AcciMap*. According to the *Australian Centre for Cognitive Work and Safety Analysis* (Australia 2003), an *AcciMap* is a multi-causal diagram, which locates various contributing factors to a collision in terms of their causal remoteness from the collision. The *AcciMap* differs from other collision analysis techniques by, not only, identifying a series of contributory factors but by assembling them into a coherent analysis that demonstrates how the factors are interrelated. By understanding the interrelationships between causal factors, it is possible to highlight the most important areas to target for intervention. No means of evaluation were found, thus no critique is given. However, Souleyrette et al. (1998) reported that the cost of implementation exceeds \$180,000 US.

Finally, there is a set of interoperable systems developed by Pd' Programming, Inc. (2003) all having the suffix name Magic—that are, *Intersection Magic*, *Map Magic*, *Project Magic*, and *Crash Magic*. These applications are all GIS-based and run in the *ArcView®* environment. These applications mainly provide crash records analysis. They

generate automated collision diagrams, pin maps of high collision locations, high collision location lists, frequency reports, and presentation graphics. Pd' Programming has been producing and distributing *Intersection Magic* for the past 14 years. According to them, this is more than twice as long as any other competitors are. Hence, it is considered, by far, the most widely used crash records analysis package of its type in the world.

2.2.3.b. Researcher-Developed Systems

FHWA has developed a collection of GIS tools for safety analysis. It was released on a CD-ROM under the name *GIS Safety Analysis Tools*. It is an *ArcView*® Project loaded with HSIS data. These tools offer various analytical techniques and procedures that all fall within the realm of the science of highway safety. However, the analysis tools could be categorised at large in two main categories: collision referencing, and collision analysis. The collision analysis tools are: *Spot/Intersection Analysis*, *Strip Analysis*, *Cluster Analysis*, *Sliding-Scale Analysis*, and *Corridor Analysis*. The *Spot/Intersection Analysis* tool is used to evaluate crashes at a user-designated spot or intersection within a given search radius. The *Strip Analysis* tool is used to study collisions along a designated length of roadway as opposed to spot or intersection. The *Cluster Analysis* tool is used to study collisions clustered around a given roadway feature such as a bridge, railway crossing, or a traffic signal. The *Sliding-Scale Analysis* tool is used to identify highway segments with a high collision occurrence. Finally yet importantly, the *Corridor Analysis* tool is used to locate high collision concentrations within a corridor. (Harkey and Griffith 1999; USA 1996)

Despite the various tools and functions that this system has, and after evaluating the system, some disadvantages were found which would undermine the analytical work done within the system. These disadvantages are:

1. The system is not easy to install and run.
2. The system has so many dependencies, which make it virtually not transportable.
3. The output reports are very difficult sometimes miraculous to interpret.
4. The underlying data structure is somehow ambiguous.
5. The system would not run unless it was installed on the root directory of the machine. This might not be suitable for systems administered by group policies that limit the usage of the root directory to system administrators.

Carreker and Bachman (2000, 215, 218) discussed the use of GIS procedures to improve speed and accuracy in locating crashes. They conducted a study of collision data in Georgia through which they have identified potential location errors. They found out that collision data are manually transliterated from paper collision reports and then located, by another operator, on a linear referencing system. As a result, data quality and quantity might hinder an operator's ability to, accurately, identify route and mile-point locations for a large percentage of collisions. Carreker and Bachman (2000) also revealed that various technical solutions—such as improved data transcription guidelines, multiple public and private highway databases, GIS, and standard relational databases potentially—could strengthen the functionality of collision location systems. Finally, they concluded that the use of GIS techniques, such as batch processing, intersection

matching, and conflation, as well as the incorporation of highway attributes from multiple sources, promised to increase the percentage of identified and located collisions.

Furthermore, they ended their conclusion with, *according to the author's point of view*, the most correct statement found in literature—that is, “The success of subsequent crash analysis system hinges on improved capabilities to locate crashes.”

In addition, Lamm, Guenther, and Choueiri (1995, 7, 13-14) analysed three safety criteria for evaluating curved highway sections including transition sections in order to address these important target areas for reducing collision frequency and severity. The analysed criteria where; achieving consistency between successive design elements, harmonizing design speed and operating speed, and providing adequate dynamic safety of driving. These criteria constitute the core of a safety module proposed for classifying highway networks or sections as good, fair, or poor designs. The analyses were done using a GIS known as SPANS⁸. The analyses was done by using discriminating colours or symbols with SPANS that resulted in separate or combined design safety levels that could be easily recognised by the safety professional. Their work laid a foundation for the use of overall safety modules. Their work was considered the first time to verify that the evaluation of highway sections or networks by an overall safety module (despite their subjective scale) is possible for design, redesign, rehabilitation, and restoration strategies.

As well, Mendoza et al. (2001, 74-75, 83) described the efforts of the *Mexican Transportation Institute* to develop a GIS-based data management system for collisions on federal highways in the various states of Mexico. This system was developed using *ArcView® GIS* after evaluating its characteristics of capacity versus price, as the resulting

⁸ SPANS is a discontinued product of PCI Geomatics, Inc.

system was intended to be delivered to each of the centers of the *Ministry of Communications and Transportation* in the states of Mexico. The system was developed by integrating into *ArcView®* a series of data for the states considered, these data included cartographic representation, classification and naming of highways, traffic volume and composition, and information on collisions. Cartography for development of this work was obtained from spatial information recorded in situ using *Global Positioning System (GPS)*. They used GPS to record geometric features of highways—that were:

1. horizontal and vertical alignments;
2. sections with different pavement types and number of lanes;
3. location of bridges and culverts; location of at-grade intersections, grade-separated intersections, and railway crossings;
4. the location of terminals and stops for passenger buses and freight trucks;
5. the location of tollbooths in toll expressways; and
6. the location of service stations.

After collecting the required data sets, segmentation and integration of the *Federal Road Classification System* was performed. Then, the next step was the integration of tabular electronic data. Finally, Mendoza et al. (2001) concluded that this system proved its success and efficiency, especially financial efficiency as the cost per km was ten dollars, including conducting GPS highway inventory. Furthermore, they recommended establishing similar systems for infrastructure and pavement management.

In addition, Quiroga and Bullock (1996, 226, 234) developed a prototype computer-based model to address the indexing needs of highway authorities—that is, indexing geometric plans, sign inventories, maintenance records, technical studies, and so forth. Their model centered on engineering report tracing but it could be extended to include other types of information. The core of their model was a procedure that linked signalised intersections and highway segments to non-spatial attribute data in a GIS environment. The indexing model included a form to facilitate further data entry into the system. They applied the system to ten-year data from East Baton Rouge Parish, Louisiana. As most of the data was linearly referenced, they deduced that geo-coding the data would be unfeasible. Therefore, they used the milepost linear referencing system to code the location of the highway features. Finally, they concluded that the framework of their model was generic and it was possible to apply it beyond the scope of *Louisiana Department of Transportation*.

On the other hand, Sadek, Bedran, and Kaysi (1999, 144, 150) used GIS as a development environment for a decision aid tool aimed to multi-criteria evaluation for route alignment. Despite the fact that their approach was away from the realm of traffic safety, their algorithm and approach could be generalised for another evaluation criteria. Furthermore, their conclusion was rightful as they stated “[...] however, in order to use the tool, a base geographically referenced model of the area of interest needs to be constructed.” Finally, they recommended constant update for the database, as GIS platforms were as good as the data underlying the application.

Saccomanno, Chong, and Nassar (1997, 18-20, 26) developed a GIS-based collision risk model for the Ontario highway network. They built their argument on the fact that highway-collision risk assessment required a thorough understanding of both the vehicle collision-involvement process and the severity of resultant injuries. Their vision was that a GIS platform was particularly suitable for this type of problem because it provided a proficient system of linking a large number of incongruent databases. Their model provided estimates of collision risk at four levels of spatial aggregation as specified by the user: network-wide, route-specific, route-section-specific, and site-specific. They used *ArcInfo*® as their development environment, with an underlying database management system. The data fed into the database management system was provided by Ontario *Highway Inventory Management System*, and the *Traffic Volume Inventory System*. Their model was limited to the Ontario 400 series highway network. They applied the model to a simple problem of black-spot identification along a given stretch of highway. Their application was limited to one year of traffic exposure. In addition, they established black-spot sections for collision involvement when observed vehicle involvement exceeded estimated vehicle involvement. A major disadvantage was that they did not account for the random nature of collisions. Another disadvantage of the whole model, was that their analytical capabilities was not extended to identify the probable causes for vehicle collision involvement.

Apart from traditional GIS-based models, Panchanathan and Faghri (1995, 91-92) developed a knowledge-based GIS system for safety analysis at rail-highway at grade crossings. Their objective was to establish guidelines and to develop an integrated system that would eventually result in collision reduction through superior access and

management of safety-related information. They developed their model by integrating a GIS application, a statistical model, and a knowledge-based expert system. Their main goal of using GIS as a development environment was to make use of its potentials to access and analyse spatially distributed data with respect to its actual spatial location overlaid on a base map of the area of interest. They were willing to benefit from using GIS spatial analysis capabilities—that are, thematic mapping, charting, network-level analysis, simultaneous access to several layers of data, and the ability to interface with external database management systems. Finally, they concluded that the use of GIS allowed better display, spatial analysis, overall better visual perception of the problem, and better data access and management. Despite their efforts to develop this system, their so-called analysis output was subjective, and did not offer much guidance to the user. In addition, their analysis criteria did not consider the randomness of collision-related attributes.

Spring (2002, 932-934, 937-938) developed a highway safety system for identification, analysis, and correction of hazardous locations. The system was developed using a knowledge-based GIS approach. He built his system to interface with existing GIS-based safety analysis systems. Spring concluded that the use of such system is feasible. Furthermore, he recommended generalisation of the knowledge-based approach to be applied within all aspects of highway engineering. On the other hand, he did not thoroughly explain his system capabilities, and his analysis criteria were vague. Furthermore, Hall, Kim, and Darter (2000, 219) provided in-depth investigation of the costs and benefits of GIS implementation. Their investigation addressed a critical need to determine the organisational impact and cost effectiveness of this enterprise-wide

information technology to achieve the maximum benefit. Their analysis resulted in an *internal rate of return*⁹ of 99.8 percent over the analysis period.

Souleyrette et al. (1998, 1, 8-19) developed a GIS-based collision analysis system under the name *GIS-based Accident Location and Analysis System (GIS-ALAS)*. The GIS-ALAS was built on several efforts by the *Iowa Department of Transportation*, law enforcement agencies, *Iowa State University*, and several other research entities to create a location referenced highway accident database for Iowa. After conducting extensive comparisons and market research, they chose *ArcView*® as their development environment. Despite their claim of GIS-ALAS analytical capabilities, it was found through literature, that its capabilities are limited to querying the underlying database, and producing graphical displays of summary data.

Finally yet importantly, the FHWA is currently developing a set of analytical tools under the name *SafetyAnalyst* for use in decision-making process to manage site-specific improvements, and to enhance highway safety by cost-effective means. These set of tools are *Network Screening Tool*, *Diagnosis Tool*, *Countermeasure Selection Tool*, *Economic Appraisal Tool*, *Priority Ranking Tool*, and *Evaluation Tool*. The *Network Screening Tool* will identify *sites with promise*. The *Diagnosis Tool* will be used to diagnose site-specific safety problems. The *Countermeasure Selection Tool* will assist users in the selection of countermeasures to reduce collision frequency and severity. The *Economic Appraisal Tool* will perform an economic evaluation of a specific

⁹ "The *internal rate of return* is the rate on the uncovered balance of the investment in a situation where the terminal balance is zero." (Riggs et al. 183)

countermeasure, or several alternative countermeasures, using the *benefit/cost ratio*¹⁰ technique. The *Priority Ranking Tool* will provide a priority ranking of sites and the corresponding proposed improvement projects, using output from the *economic appraisal tool*. Finally, the *Evaluation Tool* will provide the capability to conduct before-and-after evaluations of applied safety improvement projects. All of these tools are under development; hence, evaluating or criticising these tools was not possible. (USA 2003c)

2.3. Conclusion: The Need

As it is obvious from the literature review provided, literature is abundant on statistical and analytical safety analysis techniques. However, few GIS-based applications have been discussed in the literature. In general, all of the published work lacks some or all of the following elementary features:

1. Being Generic.
2. Having network analysis capabilities.
3. User-friendly interface and useful output.
4. Has a well-organised collision database.
5. Has a well-organised and correctly integrated means of geo-coding linearly referenced data.
6. True analytical capabilities for use in further investigation of hazardous sites.
7. Distributable and transportable system.
8. Cost effectiveness.

¹⁰ The present worth of benefits divided by the present worth of costs (Riggs et al. 1997).

To develop a robust and efficient GIS-based safety analysis system, these precautions or areas of weakness should be addressed. In addition, the proposed system should be healthy and user-friendly to survive.

Chapter 3:

Framework Development

3.1. The Vision

The proposed safety information system was named *Carleton University Safety Information system (CUSIS)*. This system has to address the users' needs discussed earlier. It has to be user-friendly, robust, efficient, and above all distributable. The concept behind this system is a vision of a GIS-based safety information system, that would enable the safety professional to scrupulously study the highway or the network and to make the appropriate conclusion or decision without having to leave his/her desk.

3.2. Framework Development and Interface Design

In order to develop the framework underlying the proposed safety information system; a study was conducted for further understanding of users' needs, expectations of potential users, and how to design an efficient and robust framework for the proposed system. According to Sinnott (1998, 375) the primary issue to be dealt with in software engineering today is re-use of software. Current software development rarely, if ever, starts from nothing. The development of software for open distributed systems is an intricate activity. Numerous issues have to be addressed to ensure that the subsystems of the system under development are correctly interoperable and capable of achieving their goals. Remoteness of software, potential for partial failure, concurrency, and language and system heterogeneity are just some of the many direct problems facing distributed systems developers.

As the proposed system will be developed using open system architecture, it is important to elaborate exactly on what is meant by an open system since numerous interpretations of this term exist in the context of distributed systems. Sinnott (1998, 375-377) reported that Leopold et al. identified several definitions that can be considered as correctly interpreting the term open as found in distributed systems literature:

- an environment and its model are easily obtained and well documented;
- the model and environment exist on many operating systems and work with many programming methodologies and languages;
- several manufacturers support and control the market; or
- the environment and its model are developed after open debating.

Within the context of this thesis, the openness of a distributed system is regarded as the extendibility of that system—that is, how new resource-sharing systems can be added to a system without disrupting existing services. Hence, the same concept used in object-oriented programming is used to develop the framework of this system. Frameworks are a natural extension of object-oriented techniques. Whilst object technology provides a basis for re-use of code, it does not provide features to capture the design experience as such. Frameworks have developed to fulfil this need. (Sinnott 1998)

A framework can be regarded as a collection of pieces of software or specification fragments that have been developed to produce software of a certain type or function. It should be noted that a framework is only partially complete. Typically, frameworks are developed so that they have holes or flexibility points in them where service specific information is to be inserted. This filling in of the holes is used to develop a massive

amount of services with slightly differing characteristics. Of course, the number of holes in a framework is directly proportional to the amount of work required to complete the service. A framework might be all but complete except for one hole, e.g., where an isolated choice of behaviours that the service can exhibit is possible. Typically, though there are likely to be numerous holes in frameworks that have to be filled. Sometimes the case is that the holes are in some way dependent upon one another. For example, a framework might have holes left to deal with costing and performance issues of the service, e.g., a higher cost means a higher throughput or picture resolution. From this, different models of framework could be classified based on the type of holes they leave. (Sinnott 1998, 377-378)

3.2.1. Frameworks Classification

According to Sinnott (1998, 378-379), frameworks could be classified in many ways. Most obviously, they could be classified based on the expected services to be generated from them, e.g., multimedia conference services, multimedia on demand services, telephony services, information systems, database management systems, information editing and processing, and so forth. Another more abstract way in which frameworks could be classified is based on their types of holes. This way is considered since it emphasises on what frameworks really are. Based upon this abstract classification, Sinnott (1998) classified frameworks into the following categories:

1. Frameworks with holes that can only be filled by certain well-defined behaviours: These behaviours are determined in advance. This is usually the case with the service framework offering several non-deterministic choices of

behaviour and the filling in of the hole corresponds to selecting one of these possible choices. This approach is very direct but unlikely to be the norm as this defies the abstract nature of frameworks. However, such an approach allows validation of the services created from frameworks to be achieved candidly.

2. Frameworks with holes based on delegating behaviour: Therefore, a hole consists of behaviour that accepts a message and redirects it to some framework specialising component. Hence, complex systems can be disaggregated into less complex subsystems, thereby abstracting unnecessary information.
3. Frameworks with holes left where the detailed modelling of the service behaviour has not been done: This type of hole requires a certain amount of engineering to be filled. This model of a hole is more likely to represent the true nature of frameworks. Several potential problems include checks that are required to ensure that the holes left are filled in correctly. The level of checking can be done in several ways depending upon how the hole itself is represented.

3.2.2. Framework Design

Literature reports numerous ways to develop frameworks. The most abundant way is to take an existing service and extract its generic features so that classes of similar services can be produced. This concept of taking a developed service and producing a framework is known as *generalisation*. The second method is *specialisation*, which is taking an abstract framework and making it less abstract. Generalisation may be noted as

capturing the main features of a given application or system in such a way that the design experience can be re-used. This can be at many levels of abstraction. For example, the cases of certain objects that involve a service are common to a collection class of similar services. Another example is the case of interfaces or operations contained within an interface are common to a class of services. On the other hand, specialisation is typically done through supplying behaviours at flexibility points. Just as generalisation can be made in several stages to result in more and more abstract frameworks, so specialisation can be done in several stages; each stage resulting in a more deterministic framework and hence a more limited set of services that can be created from the framework. Since a framework only has a finite number of holes, then each specialisation of a framework fills in one or more of these holes completely or partially. In conclusion, frameworks are tailored according to users' needs; hence, there is no specific design code for a framework. The framework designer should be committed at large to one of the two techniques commonly used—that is, generalisation or specialisation. (Sinnott 1998, 380-381)

3.2.3. User Interface Design

According to Spolsky (2001), the *user interface (UI)*, is important because it affects the feelings, the emotions, and the mood of your users. If the UI is wrong and the user feels they cannot control your software, they literally will not be happy and they would blame it on your software. If the UI is smart and things work the way the user expected them to work, they will be cheerful as they manage to accomplish small goals. Thus, a user interface is well designed when the program behaves exactly how the user thought it would.

When new users attempt to use a program, they would not come with a completely clean slate. Instead, they have some expectations of how they think the program is going to work. If they have used similar software before, they will think it is going to work like that other software. If they have used any software before, they are going to think that your software conforms to certain common conventions. They may have intelligent guesses about how the UI is going to work. This is called the *user model*, which is their mental understanding of what the program is doing for them. The program, too, has a mental model; only this one is encoded will be executed faithfully by the computer. This is called the *program model*, and if the program model corresponds to the user model, then the developer has successfully created a robust user interface. (Spolsky 2001)

The question now is how to know the users' expectations. This turns out to be relatively easy. Spolsky (2001) suggested an easy yet effective way of dealing with this dilemma. He suggested asking some potential users and briefing them on what the program does in general terms. Then the developer would describe the situation and then ask them some questions to try to guess their user model. After guessing the user model, the developer would attempt creating a program model that corresponds to it. The next step is to get the same potential users who were asked before, and this time the developer would ask them to deal with the program directly and record their experience with the software. Through this simple iterative process, a robust and simple user interface would be created. Finally, the developer has to be aware of commonly used conventions and user interfaces used within the development environment that he/she is using. This would make the program user interface consistent with the development environment.

3.3. System Analysis: Identifying User Requirements

In order to develop the robust safety information system, one must define the potential users' needs and requirements. After all, the system would be deemed successful if, and only if, it satisfied the users' requirements. The analysis method used within this thesis is object-oriented analysis¹. According to Coad and Yourdon (1991), object-oriented analysis has the following benefits:

- Improve analyst and problem domain expert interaction.
- Increase the internal consistency of analysis results.
- Explicitly represent commonality.
- Build specifications resilient to change.
- Reuse analysis results.
- Provide a consistent underlying representation for analysis and design.

Managing safety-related information and processing it is a complex problem. One can overlook the complexity of the highway system and oversimplifies it to three components—that are, the highway, the driver, and the vehicle. Other factors, however, might contribute and lead to a traffic collision. These factors include, but are not limited to, environmental conditions, visibility, and other interoperable factors. In order to override the complexity of the problem domain, the principles of object-oriented analysis were followed (Coad and Yourdon 1991). These principles are:

- Abstraction.
- Encapsulation.

¹ Object-Oriented analysis is an analysis method based upon primitive concepts—that is, objects and attributes, wholes and parts, classes and members. (Coad and Yourdon 1991, 1)

- Inheritance.
- Association.
- Communication with messages.
- Infusing methods of organisation.
- Scale.
- Categorising behaviour.

3.3.1. Abstraction

The *Dictionary of Computing* defines *Abstraction* as “the principle of ignoring those aspects of a subject that are not relevant to the current purpose in order to concentrate more fully on those that are.” Using abstraction is in other words admitting that what is being considered is complex. Rather than trying to comprehend the entire problem, only part of it is selected. This does not mean totally ignoring what was left, but it is simply not choosing to address them at this stage (Coad and Yourdon 1991, 13). Two forms of abstraction used here, procedural abstraction and data abstraction.

Procedural abstraction as defined by the *Dictionary of Computing*, as “the principle that any operation that achieves a well-defined effect can be treated by its users as a single entity, despite the fact that the operation may actually be achieved by some sequence of lower-level operations.” Procedural abstraction is used extensively by software analysts and programmers. It is simply dividing the whole system into several ordinates and subordinates. Each unit performs a function or a sub-function, according to the level of abstraction.

After reviewing the literature, and surveying existing systems, techniques, or methods that do some of the CUSIS objectives or share some functions, abstraction was done on three levels. Each level of abstraction deals with some functions and procedures that are coherent. The three levels of abstraction were named modules, applications, and functions, arranged respectively from the most to least abstract level. The first level of abstraction—that is the modules, deals with five major objectives. Those objectives were derived from the literature review and the interviews conducted with *Ontario Ministry of Transportation* officials—that are:

- Collision analysis capabilities.
- Highway characteristics analysis.
- Operational characteristics analysis.
- Consistency evaluation capabilities.
- Video logging capabilities.

Based upon these five overall technical objectives, that are derived from the goal of the system, the first level of abstraction was made. This level, the module level, organises the functionality of CUSIS into five independent modules—namely:

- *Collision Analysis Module (CAM)*: This module will provide the user with various data items and analytical tools concerning collision history over a specific highway segment, a highway, a region of the network, or the whole network. Through this module, the user would deal with various issues of collision data, either through managing the underlying database or through performing analyses based upon this database.

- Highway Characteristics Module (HCM): This module will provide the user with all possible information concerning the selected highway segment. The candidate information is related to the geometric design or the pavement design of the highway. Spatial analytical techniques will be used where applicable.
- Operational Characteristics Module (OCM): This module will provide the user with all necessary information that would give the user a comprehensive and global picture, as a metaphor, of the attributes and characteristics that affects the operation of the highway of interest.
- Consistency Evaluation Module (CEM): This module deals with one of the hottest topics in highway safety design. This module will evaluate the consistency of the design for a highway segment or an entire highway. Furthermore, it would communicate with the CAM, HCM, and OCM for queries on relevant data.
- Video View Module (VVM): This module provides video logging service to users. Basically, it will provide the user with the capability of viewing a video of a highway section, specially curves, so the highway professional, the user, can do some road safety audit tasks without having to leave his/her desk.

As for the other abstraction technique used, data abstraction, it is defined by the *Dictionary of Computing* as “the principal of defining a data type in terms of the operations that apply to objects of the type, with the constraint that the values of such objects can be modified and observed only by the use of the operations.” This technique

was applied by defining variables² and methods³ that exclusively manipulate those variables; hence, the only way to get to a variable is through a method. Variables and their methods may be treated as a fundamental whole (i.e., objects). This technique was considered in the later levels of abstraction and is to be considered through out the entire process of synthesising the system.

3.3.2. Encapsulation

Encapsulation is another technique to deal with the complexity of the safety-related problems addressed by CUSIS. The *Dictionary of Computing* defines encapsulation (see Figure 3.1) as “A principle used when developing an overall program structure, that each component of a program should encapsulate or hide a single design decision [...]. The interface to each module is defined in such a way as to reveal as little as possible about its inner workings.” Hence, encapsulation helps minimising rework when developing a new system. As the parts of the analysis effort that are most volatile were encapsulated, then the changing of user requirements becomes less of a threat to the overall effort. As the highway safety science is interrelated, and complex. Most of the attributes to be considered are susceptible to change, e.g., fifty years ago no one would have dreamed of a passenger vehicle that would go above 200 km/h, or of such safety-emphasising technologies existing nowadays in passenger vehicles. Encapsulation keeps related content together, as it minimises traffic between different parts of the work, and it separates certain specified requirements from other parts of the specification, which may use those requirements. (Coad and Yourdon 1991, 14-15)

² A *variable* is data item named by an identifier. Each variable has a type and a scope (SUN 2003).

³ A *method* is a function that serves as an interface for an object (SUN 2003).

Figure 3.1: An object featuring encapsulated variables by their methods (SUN 2003).

3.3.3. Inheritance

Inheritance is an underlying principle of the system analysis. Coad and Yourdon (1991, 15) defined inheritance as “A mechanism for expressing similarity among classes, simplifying definition of classes similar to one(s) previously defined. It portrays generalisation and specialisation, making common attributes and services explicit within a class hierarchy or lattice.” Inheritance is the technique used within CUSIS to specify common attributes and methods once, as well as specialise and extend those attributes and services into specific cases. For example, the CAM applications will inherit some common features in this module. CAM is the generalisation, and each of its applications is the specialisation. As the attributes for collision-data are the same across all applications, then there is no need to reuse them repeatedly. In addition, due to the commonality nature of the underlying database, methods are inherited from the module and propagated into the applications. This technique was applied explicitly to express commonality, beginning with the early activities of analysis.

3.3.4. Association

The Canadian Oxford Dictionary defines association as “a mental connection between ideas.” This principle was used to tie together certain things that happen at some point in time or under similar circumstances.

3.3.5. Communication with Messages

Message interaction was used within CUSIS to transfer a data value from one object to another. This was done to manage the complexity of data. The message interaction is conducted through the interfaces (i.e., between methods not between the variables directly). This guarantees data integrity and consistency across all objects in the system.

3.3.6. Infusing Methods of Organisation

According to *Occam's razor*, the principle of parsimony, one should not make more assumptions than the minimum needed. In order to minimise the number of assumptions made within the analysis, the following steps to organise the analysis effort were adopted:

1. The experience derived from literature was differentiated into particular objects and their attributes (i.e., variables).
2. Whole objects were differentiated from their component parts.
3. Different classes of objects were formed and differentiated.

3.3.7. Scale

The domain of safety information management is very large. One could be overwhelmed, which would lead to misconception of the whole problem. The principle of scale helps relating to something very large without being overwhelmed, thus harmonising the relationship of the termed parts of the system within the analysis, and the users' perspective. With applying the concept of scale, analysis notation and strategy can include ways to guide the user through a larger level.

3.3.8. Categorising Behaviour

Simply put, object behaviour is its methods. In other words, behaviour is the interface through which an object exchanges messages and communicates with other objects. The most familiar categories of behaviour are those of a human being—that are based on immediate causation, on similarity of evolutionary history, or on the similarity of function (Coad and Yourdon 1991, 17). The applicable category to software development is the first category, which implies deploying an event-response strategy throughout the objects. This was done by, whenever possible, modeling the methods of the objects to respond to the user input events, or to the events caused by other objects.

3.4. Framework Architecture

In order to make CUSIS expandable, an object-oriented design⁴ was adopted. As in object-oriented programming, the building unit of the software is an object. An object could be defined as a software bundle of related variables and methods. After performing the analysis as described earlier, the whole-part relationships were realised and were done on three abstraction levels (see Figure 3.2). As mentioned earlier the three levels of abstraction were system modules, system applications, and system functions respectively arranged from the most to least abstract.

⁴ “*Object-oriented design* is a software design method that models the characteristics of abstract or real objects using classes and objects.” (SUN 2003)

Figure 3.2: CUSIS whole-part structure notation

In addition, functional decomposition strategy was used in achieving the structure. As each *CUSIS-Module* performs a certain function, each *CUSIS-Application* performs a certain sub-function, and each *CUSIS-Function* serves as a functional interface.

3.5. Framework Components and their Functions

As described earlier the CUSIS was functionally decomposed into five modules. These modules (as described in section 3.3.1) were *Collision Analysis Module*, *Highway Characteristics Module*, *Operational Characteristics Module*, *Consistency Evaluation Module*, and *Video View Module* (see Figure 3.3). In this section, each module is described and its child applications are illustrated with their functions.

Figure 3.3: CUSIS modular structure.

3.5.1. Collision Analysis Module (CAM)

The majority of the analytical methods to investigate highway safety are dependant upon historic collision data. In addition, it is very useful to outlook a wide range of historic data, to make a sound judgement. Being able to, graphically, view these data is the optimum way to outlook the *big picture*. At large, this CUSIS-Module would serve as a database management facility for collision data. It would provide querying tools and data management tools. As mentioned earlier each CUSIS Module consists of several CUSIS Applications, for CAM these CUSIS-Applications are (see Figure 3.4):

- Collision Database: This CUSIS-Application provides collision-database management capabilities. The CUSIS-Functions within the Collision Database application would be a function for plotting collision records, a function to query collision database, a function to add a new collision record, and another for deleting a collision record.
- Collision Analysis: This CUSIS-Application would provide the user with an analytical tool that compromises collision prediction and collision history with the employment of an Empirical Bayes technique. This CUSIS-Application constitutes of two CUSIS-Functions, namely: Intersection Analysis, and Highway Segment Analysis. The first of the two would provide the analytical tool for highway intersections or interchanges. While the second one, would provide the analytical tool for highway segments.
- Hazardous Locations: This CUSIS-Application would employ an Empirical Bayesian technique to identify collision prone locations with the aid of a knowledge base. Using the knowledge base as an underlying foundation for

this CUSIS-Application would compromise for missing or erroneous data in the collision history. In addition, it would be useful for evaluating new highway designs that have no history yet. This CUSIS-Application is clearly structured, and developing it would require a breakthrough in highway safety science.

Figure 3.4: Collision Analysis Module structure

3.5.2. Highway Characteristics Module (HCM)

As it is evident from the literature review, highway geometric design features dramatically influence the highway safety. In order to take the correct and feasible decision, the highway safety professional has to be fully aware of the various highway characteristics. This CUSIS-Module would provide the highway safety professional with all possible information concerning a selected highway segment. This information is related to the geometric design or the pavement design of the highway. As reported earlier “[...] A picture worth thousand words”. Hence, representing the highway characteristics data graphically would be convenient and suitable for this kind of

problem. This CUSIS-Module consists of two CUSIS-Applications (see Figure 3.5), they are:

- Alignment Data: This CUSIS-Application will provide the highway geometric design data for a user-selected highway segment. It consists of three CUSIS-Functions: The first one is for displaying vertical alignment data. The second one is for displaying horizontal alignment data. Finally, the third one is for displaying cross-section data. The display is prepared online by querying the spatial and tabular databases underlying this CUSIS-Application.
- Pavement Characteristics: This CUSIS-Application would complement its twin, the Alignment Data application. It consists of two CUSIS-Functions: The first one will provide structural data for the highway segment selected. The candidate data would be the design method used and the as built design parameters used. The second one will provide surface characteristics, which are mainly lateral and side friction indices, and pavement serviceability indices. This CUSIS-Application would have an online connection with the national pavement database of the Canadian Long-Term Pavement Performance Program, *C-LTPP*.

Figure 3.5: Highway Characteristics Module structure

3.5.3. Operational Characteristics Module (OCM)

Plausibly, highway traffic collisions happen only when there are vehicles being operated on the highway. In addition, it was evident from the literature review that the operational characteristics of the highway segment, highway, or the highway network noticeably influence the collision patterns. The overall objective of this CUSIS-Module is to provide the user, highway professional, with a comprehensive set of data that would describe the operational characteristics of the highway segment of interest. Various operational data would be provided to the user through this CUSIS-Module. This CUSIS-Module would consist of three CUSIS-Applications (see Figure 3.6), each to provide a different set of operational data. These CUSIS-Applications are:

- Traffic Volume: This CUSIS-Application would display the traffic volume data, in terms of Annual Average Daily Traffic, AADT. The user will select a highway segment and then a year, or average across all years, to display the corresponding AADT figure. The AADT data are retrieved from a database

that has AADT data for several years. Initially a ten-year period would be considered.

- Speed: This CUSIS-Application would provide the user with various data entries for speed on a highway segment. This CUSIS-Application consists of three CUSIS-Functions. The first CUSIS-Function is responsible for providing the user with operating speed estimates for the highway segment of choice. The user would have the ability to choose between several methods of calculation to display an operating speed profile. The second CUSIS-Function would provide the user with the design speed profile for the chosen highway segment. The third CUSIS-Function would provide the user with the actual running speed along a selected highway segment. This would be useful indicator for economic appraisals.
- Travel Advisories: This CUSIS-Application would be an online application connected to a data acquisition system and a database to collect the surface conditions (resulting from weather conditions), visibility measures, and road closures of every highway segment considered within the CUSIS.

Figure 3.6: Operational Characteristics Module structure

3.5.4. Consistency Evaluation Module (CEM)

Highway design consistency is one of the major factors that contribute to the highway safety evaluation. As the highway user builds expectations while driving along a highway, he would assume that whatever is lying ahead would be the same as that passed. The overall objective is to ensure that successive geometric design features produce a harmonised driving experience with no surprise events. Hence, critical manoeuvres could be avoided and safe highway operation could be achieved. This CUSIS-Module consists of five CUSIS-Applications that check five criteria of design consistency (Figure 3.7). These CUSIS-Applications are:

- Operating Speed: This CUSIS-Application provides means of evaluating geometric design consistency based on operating speed. Operating speed is used widely as the most abundant measure for geometric design consistency evaluation. This CUSIS-Application will allow the user to choose the values of operating speed through querying the OCM, and giving the option for the user to choose an estimate speed or actual speed if present. This CUSIS-Application consists of two CUSIS-Functions. The first CUSIS-Function will evaluate the geometric design consistency based on operating speed and using differential values of operating speed and design speed (i.e., Safety Criterion I). The second CUSIS-Function will evaluate the geometric design consistency based on operating speed using operating speed deferential (i.e., Safety Criterion II). (Lamm et. al. 1995)
- Side Friction: As it is evident from the literature (Hassan 2002) when a vehicle experiences excessive centripetal forces, rollover or head-on collisions

are imminent. One of the major factors that contribute to these centripetal forces is the side friction between the vehicle tires and the pavement surface. This CUSIS-Application will evaluate the geometric design consistency based upon the side friction demand and side friction supply (i.e., Safety Criterion III). (Lamm et al. 1995)

- *Alignment Indices:* This CUSIS-Application will provide a tool for evaluating geometric design consistency based upon Alignment indices. Alignment indices have the privilege of quantifying the general character of a geometric design element in a highway section. Candidate alignment indices for horizontal elements are curvature change rate, degree of curvature, curve length, average radius, and average tangent length (Hassan 2002). In addition, candidate alignment indices for vertical elements are vertical curvature change rate, average rate of vertical curvature, and average gradient. On the other hand, there is one candidate composite index, which is combination curvature change rate. This CUSIS-Application will output a table with the rating of a highway segment or a highway according to each alignment index, however, a suitable rating system is not yet available⁵. The evaluation will be based upon differential values of the indices between successive design elements.
- *Driver Workload:* Literature indicates that driver workload is a highway-user based measure for geometric design consistency. This CUSIS-Application will evaluate the geometric design consistency based upon driver workload. It will utilise an underlying knowledge base for driver workload measures. This

⁵ Research is undergoing at Carleton University by Miss M. Awatta, to address design consistency evaluation based upon alignment indices.

knowledge base will have expert ratings for geometric design features, and the experience of drivers collected from a field test. The evaluation would be done by using the differential between driver workload supply and driver workload demand, however, quantifying driver workload is not well established. Therefore, developing this CUSIS-Application would require an established method for quantifying driver workload.

- *Collision Prediction:* In the same manner discussed above, the driver would build an expectation of predicted collisions over a highway segment or a highway. The driver's expectation should be consistent and harmonised to avoid hazardous manoeuvres. This CUSIS-Application consists of two CUSIS-Functions that will evaluate the design consistency of the highway. The first CUSIS-Function will evaluate the design consistency using the differential predicted collision frequency between successive highway elements. The second CUSIS-Function will evaluate design consistency using the differential between predicted and actual collision frequency for the same highway element. Then based upon the degree of discrepancy, each highway element will be tagged with a subjective rating. It should be mentioned that this topic is an open research topic. An established and recognised method of evaluation does not exist yet.

Figure 3.7: Consistency Evaluation Module structure

3.5.5. Video View Module (VVM)

This module consists of two CUSIS-Applications (see Figure 3.8). The first CUSIS Application connects to a video library, and displays the selected video for a driving experience on a highway design element. The second provides a more flexible tool, where some effects could be introduced to the actual video to simulate some condition on the highway element. These applications could benefit researchers or safety auditors so they can have a real-time driving experience over a highway segment without having to leave their desk.

Figure 3.8: Video View Module structure

3.6. Choosing the GIS Environment

A GIS environment had to be selected for synthesising this framework. The proposed GIS environment had to provide certain capabilities and features:

- Provide means of customisation through an object-oriented programming language.
- Allow the developed system to be transportable, modular, and distributable.
- Affordable and have a well-established market presence.
- Easy to use and have a short learning curve (at the user side).
- Supports various data formats, including GIS data, database files, image files, and so forth.

A market survey was conducted, along with reviewing users' reviews and opinions, to find a GIS environment that provide these capabilities and features. The market survey concluded that *ArcView® GIS* version 3.x is the most feasible option. *ArcView® GIS* meets all of the above requirements, and it could be said that it dominates the worldwide market for GIS. *ArcView® GIS* version 3.x is a product of Environmental Systems Research Institute, Inc., with more than 500,000 copies sold worldwide (ESRI 1996a). In addition, the majority of transportation jurisdictions use *ArcView® GIS* as their GIS environment. *ArcView® GIS* could be customised for a certain specialisation, in our case it is CUSIS, using the built-in *Avenue* object-oriented scripting language. Furthermore, these customisations could be wrapped up into a distributable extension. In addition, within *ArcView® GIS* data from virtually any source could be integrated and

utilised in analyses, hence, *ArcView® GIS* was selected as the GIS development environment for synthesising CUSIS.

Chapter 4:

Data Collection, Formatting, and Processing

4.1. Introduction

After reviewing the literature, identifying user needs, and developing the framework, data have to be collected to synthesise the CUSIS. However, the underlying structure of the candidate data should be studied first. This would help to understand the referencing systems and methods used within the collected data domain. The next step is to identify candidate data, and collect them. Follows is data reformatting such that it would have a consistent and uniform organisation. This would be done by unifying the coding systems used through one unique system. The last step is data processing to trim unwanted data items, prepare the data for the synthesis phase, and to make sure that the data is consistent. Data consistency is a crucial measure of its integrity, as it would eliminate potential errors, thus ensuring robust processing on the application side.

4.2. Highway Data Referencing Systems

Highways are linear features by nature. Highways are considered linear, as their width is relatively negligible when compared to their length. Hence, traditional methods and systems used to locate highway features relied upon linear referencing techniques. Hence, linear referencing systems¹ and methods² were abundantly used within traditional safety analysis techniques and methods. Although *Global Positioning Systems (GPS)* are versatile and common nowadays, yet the majority of highway jurisdictions maintain their

¹ Linear Reference Systems are defines as “the total set of procedures for determining and retaining a record of specific points along a [highway]. The system includes the location referencing method(s), together with the procedures for storing, maintaining, and retrieving location information about points and segments on the highways.” (Smith, Harkey, and Harris 2001, 11)

² Linear Reference Methods are defined as “the technique used to identify a specific point (location) or segment of highway, either in the field or in the office.” (Smith, Harkey, and Harris 2001, 11)

highway data and inventory in a linearly referenced system. Smith, Harkey, and Harris (2001) categorised the most common location referencing methods into five categories:

- Route Milepost System: This system is the most commonly used linear referencing system in the United States of America at the State level. Sometimes in US literature, it is referred to as “Route Mileage” system. Within this system, locations are referenced from the start of the highway or the jurisdiction boundary with the distance measured in miles and rounded to the nearest hundredth of a mile (see Figure 4.1). The main disadvantage of this system is the change of length due to realignment of the highway. (Smith, Harkey, and Harris 2001, 12)

Figure 4.1: Route Milepost System (Smith, Harkey, and Harris 2001, fig. 5).

- Route-Reference Post System: This system depends upon fixed reference points (i.e., posts) along the highway (see Figure 4.2). The fixed points are also known as *reference posts*. A location is referenced by a distance and direction. These reference posts are fixed and do not move with realignment of the highway. Thus, locations are not dependant upon the highway length or alignment. (Smith, Harkey, and Harris 2001, 12)

Figure 4.2: Route-Reference Post system (Smith, Harkey, and Harris 2001, fig. 6).

- Link-Node System: This system divides the highway network to nodes and links (see Figure 4.3). Nodes are usually physical features, such as intersections, or interchanges. Nodes are considered unique, and are given a unique identifier within the highway network. Links are the logical connection between nodes. Links are referred to by the beginning and ending nodes. Each link is given a unique identifier, which is usually derived from the node identifiers. When referencing a location an offset along the link is given from the beginning node. Usually beginning nodes have the nearest or lowest number. This is the system used within the *Province of Ontario*. (Smith, Harkey, and Harris 2001, 12)

Figure 4.3: Link-Node system (Smith, Harkey, and Harris 2001, fig. 7).

- Route-Street Reference System: This system is abundant in United States municipalities. This system relies on the local street system. Within this system, a location on a certain street is referred to, by one of two ways. The first way is using a distance and direction from a reference street. The second way, which is less precise, is by using two reference streets and no distance measured. (Smith, Harkey, and Harris 2001, 13)
- Geographic Coordinate System: Finally yet importantly, this system of referencing locations does not depend upon a local feature. It depends upon geographic coordinates obtained by a GPS. This system has a major advantage

over other referencing systems, as it would report the linear nature of a traffic collision. Traffic collisions start at one point and end on another point. The reporting personnel would utilise this system to report the pathway of the collision. Other referencing systems refer to collisions as point events. Furthermore, digitally processing this data is much easier, and it minimises the human error within other referencing systems. (Leore 2003; Smith, Harkey, and Harris 2001, 13)

The *Ontario Ministry of Transportation* utilises the Link-Node linear referencing system for reporting collision data, and other highway features. A node is an intersection between two roadways, or a municipality boundary, which is given a unique five-digit number. Links, or sections, are logical connections between these nodes, where the beginning node is the one with the lower number. The node numbers are referred to as *Linear Highway Referencing System (LHRS)* number. When noting a link, or a highway section, it is referred to by the beginning LHRS number. The unit of measuring link lengths is kilometres rounded to the nearest hundredth of a kilometre. (Boutilier 2002; Edwards 2002)

4.3. Data Collection

Since the proposed system is GIS-based, the first data item to be collected was a GIS map of Ontario highways. The *Ontario Ministry of Transportation (MTO)* has created a *Digital Cartographic Reference Base (DCRB)* of Ontario at a data capture map scale of 1:100,000. Data were captured by digitising from the MTO County Map Series for Southern Ontario and from the *Provincial Map Series* produced by the *Ministry of*

Natural Resources at a data capture scale of 1:100,000. The original data capture was in MicroStation® Version 4 format and then converted into ARC/INFO format by MTO. MTO claims that all errors in the data were corrected in-house and the data inserted into the DCRB. However, in the DCRB release notes MTO stated, “The current data set is in transition and will undergo structural, format, and data content changes that are noted in an accompanying addendum.” The DCRB data are presented in ArcInfo® coverage³. All data are in geographic coordinates (i.e., latitudes and longitudes). Horizontal datum for all data is the *North American Datum 1983*⁴ (NAD83). The horizontal positional accuracy of the DCRB data is ± 100 metres. However, the LHRS point theme⁵ was missing from the DCRB version 3.0, so it was taken from DCRB version 2.0, which used the *North American Datum 1927*⁴ (NAD27).

The second data item to collect was a set of collision data tables. MTO maintains an information system named “Accident Information System” for storing and managing collision data. These tables were acquired from MTO eastern region in Microsoft® Excel format. Each table had the collision history for a specific highway from, and including, the year 1995 to, and including, the year 1999. The highways included were highways 7, 15, 17, and 41. Each record in the tables represented a vehicle that was involved in a collision (e.g., a three-vehicle collision would be stated in three data records). The following table (Table 4.1) illustrates the data fields present in the data tables as provided by MTO.

³ ArcInfo coverage is a GIS data format created by Environmental Systems Research Institute, Inc. (Patterson 2002, 14)

⁴ For more information on NAD83 and NAD27 geodetic datum see (Patterson 2002; Dana 2003).

⁵ A theme is a set of geographic data. (Patterson 2002, 13)

Table 4.1: Collision-data tables field description

Field number	Field name	Description	Data type
1	LHRS	LHRS number used in referencing the collision location	Integer
2	Offset	Offset from the LHRS point towards the higher LHRS number, in kilometres and rounded to the nearest hundredth of kilometre.	Real
3	Microfilm	Microfilm number for the police collision report	Integer
4	Yr-Mo-Dy	Date of collision	Date
5	Time	Time of collision	Integer
6	Day	Day of week, shortened in a three-letter form.	String
7	Class	Collision class (Fatal, Injury, or PDOOnly)	String
8	Veh#	Vehicle ID	String
9	Mainline	A variable describing whether the vehicle was on the mainline road or on the minor junction. This field is useful for collisions happening at intersections.	String
10	Initial Impact Type	The vehicle initial impact type.	String
11	Light Condition	The light conditions at time of collision	String
12	Road Location	The road location of the vehicle involved in the collision.	String
13	Alignment Road 1	A description of the road alignment	String
14	Environment Condition 1	Represents the weather condition at time of collision.	String
15	Surface Condition Road 1	Represents the road surface condition, due to weather factors.	String
16	Driver Action	Describes the driver action at time of collision.	String
17	Driver Cond	Describes the driver condition at time of collision.	String
18	Veh Type	Vehicle type for the vehicle concerned in this data record.	String
19	Veh Manoeuvre	Vehicle manoeuvre at time of initial impact.	String
20	Event 1	A description of the first event right after initial impact until vehicle came to rest or until second event.	String

21	Event 2	A description of the second event right after the first event until vehicle came to rest or until third event.	String
22	Event 3	A description of third event right after the second event.	String

The third data item to be collected was the traffic data. MTO maintains an information system under the name “Traffic Volume Information System” for traffic data. The data were received from MTO in a portable document format⁶, *PDF*. The report contained traffic volume data for the Ontario highway network covering a ten-year period from 1991 to 2000. The traffic volume data reported represents traffic volume over a highway segment and referenced by the beginning LHRs number of this highway segment. The traffic volume data-report represented the seasonal variation in traffic volumes by reporting traffic volumes in four categories as follows:

- “AADT: Annual Average Daily Traffic; defined as the average twenty four hour, two way traffic for the period January 1 to December 31.
- SADT: Summer Average Daily Traffic; defined as the average twenty four hour, two way traffic for the period July 1 to August 31, including weekends.
- SAWDT: Summer Average Weekday Traffic; defined as the average twenty four hour, two way weekday traffic for the period July 1 to August 31, excluding weekends.
- WADT: Winter Average Daily Traffic; defined as the average twenty four hour, two way traffic for the period January 1 to March 31, plus December 1 to December 31, including weekends.” (Province of Ontario 2000)

⁶ PDF, Portable Document Format. Created by Adobe Systems, Inc.

The fourth data item was the description of each LHRS point used in GIS maps, referencing collision data, and referencing traffic data. Two PDF files were received from MTO eastern region, containing the required description. The first file had the highway key-points for collision location identification in highway sequence. The second file had the composite LHRS referenced features listing, which are more detailed and are less aggregated compared to the first file.

The fifth data item was horizontal and vertical alignments for the four highways of interest (i.e., Highways 7, 15, 17, and 41). Not all of the alignment of these highways was acquired (see Table 4.2). For Highway 7, the alignment acquired covered only that portion passing through Lanark County. For Highway 15, an electronic version of the alignment covering the entire section of the highway starting from its intersection with Highway 7, up to its intersection with Highway 43. For Highway 17, the alignment acquired covered that portion of the highway passing through the Regional Municipality of Ottawa-Carleton, and Renfrew County. Finally, for Highway 41, the alignment acquired covered that portion of the highway passing through Renfrew County.

Table 4.2: Highway alignments received from MTO

Highway number	County	Township	Start Station	Section Length
7	Lanark	Beckwith	10+000.000	10544.055 m
7	Lanark	Ramsy	10+000.000	7385.684 m
15	Lanark	Beckwith	--	--
15	Lanark	Montague	--	--
17	RMOC	Fitzroy	10+000.000	16061.242 m
17	Renfrew	Horton	10+000.000	16840.763 m
17	Renfrew	McNab	10+000.000	21761.193 m
41	Renfrew	Grattan	100+11.71	455.990 ft
41	Renfrew	Alice, Stafford, and Pembroke	10+000.000	10200.000 m

Finally yet importantly, the sixth data item that had to be collected was the video clips for the highway sections. Sixteen horizontal curves were selected for videotaping. Then the videos were shot using a digital video camera from the passenger's eye height while driving just under the posted speed limit of these curves. Then the videotape was transferred to the computer using Microsoft® Windows Movie Maker and Ulead® VideoStudio™ version 4, in MPEG-2⁷ format. Each curve was taped twice, with each time in a direction. Maximum effort was done to shoot all the videos within the same time of day to eliminate discrepancies in the video log. Those sixteen curves were selected to represent the majority of the horizontal curves, at large, of the rural highway network in Ontario. The following table (Table 4.3) shows the description of the sixteen curves that were videotaped.

Table 4.3: Description of videotaped horizontal curves

Curve number	Highway number	County	Township	Station ⁸
1	31	Stormont, Dundas, and Glengarry	Mountain	27+731.038
2	31	Stormont, Dundas, and Glengarry	Mountain	16+551.184
3	31	Stormont, Dundas, and Glengarry	Williamsburgh	621+53.91
4	31	Stormont, Dundas, and Glengarry	Williamsburgh	583+91.94
5	43	Stormont, Dundas, and Glengarry	Mountain	115+82.16
6	43	Stormont, Dundas, and Glengarry	Mountain	220+46.43
7	43	Stormont, Dundas, and Glengarry	Mountain	254+49.29

⁷ MPEG-2 refers to the standard ISO/IEC – 11172, for coding audio-visual information to a digital compressed format.

⁸ Highway station number for the horizontal point of intersection of the curve.

8	43	Stormont, Dundas, and Glengarry	Mountain	262+51.60
9	12	Stormont, Dundas, and Glengarry	Finch	38+51.72
10	12	Stormont, Dundas, and Glengarry	Finch	79+33.66
11	12	Stormont, Dundas, and Glengarry	Finch	286+90.81
12	12	Stormont, Dundas, and Glengarry	Finch	305+58.89
13	41	Renfrew	Grattan	11+120.197
14	41	Renfrew	Grattan	10+816.406
15	41	Renfrew	Grattan	382+00.02
16	41	Renfrew	Grattan	397+06.63

4.4. Data Reformatting

Data items received from MTO could not be directly used within a GIS environment unless it was reformatted and restructured so it would suit the potentials of the GIS environment being used. Quality data are the essential underlying foundation of any information system. Hence, data consistency and integrity had to be revised, to eliminate any errors or discrepancies in the data. Through data reformatting and then processing, consistency and integrity were achieved.

The first data that had to be reformatted was the underlying GIS map covering the scope of CUSIS. Five Arc/Info coverages and a Shapefile⁹ from the DCRB version 3.0 and 2.0, respectively, were used as the underlying map for CUSIS. A view¹⁰ was created in ArcView® GIS and the five Arc/Info coverages and the Shapefile were added as themes. These five Arc/Info coverages and the Shapefile were arranged such that

⁹ "The Shapefile is ArcView's native data format." (Patterson 2002, 14)

¹⁰ "A view displays a set of cartographic data referred to as themes." (Patterson 2002, 13)

coverage or Shapfile is in a layer, and the layers are arranged one above the other (see Figure 4.4). The Arc/Info coverages had the datum NAD83, while the Shapefile had the datum NAD27. The next step was to convert the Arc/Info coverages to Shapefiles. This was done using *ArcView® GIS* native converting feature. Then the Arc/Info coverages were removed from the view. The Arc/Info coverages and Shapefile arrangement from the lower most themes to the top most themes is:

- “Utmzones”: This Arc/Info coverage has polygons showing the Universal Transverse Mercator, *UTM*, zones¹¹ within Ontario.
- “Upper_tier”: This Arc/Info coverage has polygons defining the boundaries of Counties within Ontario.
- “Mto_geo_twp”: This Arc/Info coverage has polygons defining MTO Geographic townships boundaries.
- “Roads”: This Arc/Info coverage has lines that represent Ontario highway network.
- “Roads”: This Arc/Info coverage has points representing nodes of the Ontario highway network.
- “lhrrs.shp”: This Shapefile has points representing the LHRS points within Ontario.

¹¹ “The Universal Transverse Mercator map projection and its associated referencing system provide a linear system of spatial referencing in a spherical world. [...] The UTM system is used extensively in North America [...]. The UTM system subdivides the earth into 6° longitudinal by 8° latitudinal zones.” (Patterson 2002, 3)

Figure 4.4: A screen shot showing a view created and the Arc/Info coverages added to ArcView® GIS environment.

The second data item to reformat was the collision data files. The original format was Microsoft® Excel format. The target format was dbaseIV¹² format. This format conversion was done using Microsoft® Excel. The field names within ArcView® GIS are limited to ten characters with no spaces. Therefore, the field names were changed as illustrated in the following table (Table 4.4):

¹² dbaseIV format is a file format for dbase® database management system produced by dbase Inc. This format is used in ArcView GIS® for storage and management of data tables.

Table 4.4: A table showing field name change for collision data tables

Original field name	Modified field name
LHRS	LHRS
Offset	Offset
Microfilm	Microfilm
Yr-Mo-Dy	Date
Time	Time
Day	Day
Class	Class
Veh#	Veh#
Mainline	Mainline
Initial Impact Type	Initial_im
Light Condition	Light_cond
Road Location	Road_locat
Alignment Road 1	Alignment
Environment Condition 1	Environmen
Surface Condition Road 1	Surface_co
Driver Action	Driver_act
Driver Cond	Driver_con
Veh Type	Veh_type
Veh Manoeuvre	Veh_manoeu
Event 1	Event_1
Event 2	Event_2
Event 3	Event_3

The date in the original files was formatted as yy/mm/dd, i.e., two digits representing the year followed by a slash, then two digits representing the month followed by a slash, then two digits representing the day of the month. The Date field was reformatted in the modified files to be yyyyymmdd (i.e., four digits representing the year, the subsequent two digits represent the month, and the final two digits represent the day of the month).

The third data item to be reformatted was the traffic volume data. A table was created in dbaseIV format to record this data. According to the scope of CUSIS, only

AADT data will be utilised. Hence, the table data fields were structured as shown in the following table (Table 4.5):

Table 4.5: Traffic volume data-table structure.

Field number	Field Name	Description	Data type
1	Hwy	Highway number	Integer
2	Lhrs	Section starting LHRs number	Integer
3	Sec_len	Section length as reported in the traffic volume data file, in kilometres and rounded to the nearest hundredth of a kilometre	Real
4 .. 13	1991 .. 2000	AADT in that Year	Integer
14	Average	Average AADT from 1991 to 2000	Integer

A data record would represent the AADT for each highway section from the year 1991 to the year 2000, with the last field in the record as the average.

Finally yet importantly, the fourth data item to be reformatted is the alignment data¹³. This was done by creating six tables in dbaseIV format. The first five tables had the elevation data for all highway segments. The tables were for centerline, Northbound, Southbound, Eastbound, and Westbound lanes respectively. The sixth table had the horizontal curves data for all highway segments. The table structure is illustrated in the following two tables (Table 4.6 and Table 4.7):

¹³ In practice, this step was done after processing the data files and integrating them into GIS, as it requires cross-referencing highway segments between GIS map and acquired alignment.

Table 4.6: Table structure for elevation data.

Field number	Field Name	Description	Data type
1	County	County name	String
2	Geotwnshp	MTO geographic township name	String
3	Lhrs	Section start LHRS point	Integer
4	Offset_m	Offset in meters from LHRS point to the elevation point	Real
5	Station	Station number for elevation point	String
6	Elev_m	Elevation in meters for the elevation point	Real

Table 4.7: Table structure for horizontal curves data.

Field number	Field Name	Description	Data type
1	County	County name.	String
2	Geotwnshp	MTO geographic township name.	String
3	Lhrs	Section start LHRS number.	Integer
4	Offset_m	Offset in metre from LHRS point to the point of interest, rounded to the nearest millimetre.	Real
5	Station	Station number for the point of interest.	String
6	Inv_r	10^5 divided by curve radius in meters.	Real
7	Rad	Curve radius in meters rounded to the nearest millimetre. The radius of the tangent is reported as "INF", for infinity.	String
8	Dir	Direction of horizontal curvature when driving in the direction of lower LHRS number to higher LHRS number. L: Left, R: Right.	String

9	Type	Type of point being recorded. START: First point recorded in the highway, END: Last point recorded in the highway, BC: Beginning of horizontal curvature for circular curves not proceeded with transition curves, EC: End of horizontal curvature for circular curves not proceeded with transition curves, TS: Point where tangent ends and spiral curve starts, SC: point where spiral curve ends and circular curve starts, CS: point where circular curve ends and spiral curve starts, ST: point where spiral curve ends and tangent starts.	String
10	Sn	Serial number of this curve within all curves on highway. Curves are numbered in the direction of Lower LHRS to higher LHRS.	Integer

4.5. Integrating Data into ArcView® GIS

It is estimated that eighty percent of all data contain a geographic component such as country, state, ZIP Code, or street address. By using the power of ArcView® GIS, these data could be integrated for analysis and better decision-making (ESRI 2003). After reformatting and restructuring the acquired data files, they had to be integrated within the *ArcView® GIS* for processing. In a standalone configuration, *ArcView® GIS* supports integration of database tables in dbaseIV file format. Furthermore, *ArcView® GIS* offers many tools for analysing tabular data, and cross-referencing with geographic data. The various analytical capabilities include querying, sorting records, summarising records, modifying table properties and assigning aliases to field names, automatically calculating values for a field, editing a record, adding or deleting a field, data reclassification,

combining tables, building relations between tables (i.e., one-to-one, one-to-many) (Patterson 2002, 58-76).

4.6. Data Processing

Because the data were collected from various resources, problems have arisen from aligning and overlaying the data items. Thus, some data had to be chosen to be the underlying foundation for all other data items. The node-link system representing the Ontario highway network was chosen for this task, because logically data would be related to a highway segment, but not the other way around. The first task that had to be performed was setting the data projection for the area of interest. Ontario lies in four UTM zones, namely UTM zone 15, 16, 17, and 18. The view projection was set to UTM zone 18, because the scope of CUSIS at this phase is limited to Eastern Ontario region.

The LHRS Shapefile originally had a NAD27 datum, while all of the other Shapefiles used the NAD83 datum. The NAD83 datum is more updated and accurate; hence, it should be used for developing CUSIS. This difference in datum resulted in a planar shift in the order of 200 to 400 meters for every LHRS point. Several feature manipulation engines were tested to translate the LHRS Shapfile datum from NAD27 to NAD83. Finally, FME Suite¹⁴ gave promising results during the trials, and was used to translate the datum from NAD27 to NAD83.

Since the data were in decimal degrees, and the ready-made software package, FME Suite, uses only four decimal places, there was still a planar shift in the order of three to six metres for every LHRS point. Fine-tuning these datum translations was done

¹⁴ FME Suite is a product of Safe Software.

by using a script that was specially developed by the author for this task, using Avenue. This script searched for the nearest node within three meters from every LHRS point. The script was designed that if more than one node were found, the script prompted for manual identification, however, this case did not happen and the node numbers were unique to LHRS points. Then the script copied the geographic attributes of the node to that of the LHRS point, thus accurately and precisely locating the two points over each other. Only one LHRS point was not in sequence of logical order. A site visit was conducted to the place where it was supposed to exist. Then by comparing the geometric features description, the correct location was identified and that point was moved to it. Then the LHRS points for the four highways of interest were separated to enhance data visualisation into a new theme, named "Data LHRS."

After finishing the required adjustments for the LHRS points, highway links had to be revised to check their segmentation. Each highway segment had to match the description provided by MTO; therefore, between any pair of successive LHRS points there had to be a unique link. This procedure is very expensive in terms of computing time, thus, it was conducted only for those segments for the four highways of interest, and in the same time existed in MTO eastern region. A procedure known as combining features was used to combine line segments together. In addition, there was no need to split line segments.

Within ArcView® GIS, one can use the clip operation to cut out a piece of one theme using another theme as a "cookie cutter" (Patterson 2002). Using this operation, the county boundaries from the counties theme was used to extract the Ontario roads

from the roads theme, and to create a new theme containing a smaller number of roads. Then the four highways of interest were separated for enhanced visualisation into a new theme, named “Data HWYs.” The attributes for this new theme were edited and fields were recoded in order to enhance and optimise CUSIS synthesis. The attributes used for this theme are illustrated in the following table (Table 4.8):

Table 4.8: Attributes of “Data HWYs” theme.

Field number	Field Name	Description	Data Type
1	Shape	A field for storing geographic information for the feature.	Native
2	Segment	Highway segment number. The three right most digits are the segment serial number within the highway, and the rest of the digits are the highway number.	Integer
3	Ref_LHRS	The start LHRS according to MTO data files.	Integer
4	Sec_Len	Reported length in kilometres, rounded to the nearest hundredth of a kilometre, of the highway section according to MTO data files.	Real
5	Length_Km ¹⁵	Reported length in kilometres, rounded to the nearest hundredth of a kilometre, of the highway section according to the GIS map.	Real
6	F_LHRS	LHRS number for the start-point in direction of digitisation.	Integer
7	T_LHRS	LHRS number for the end-point in direction of digitisation.	Integer
8	Hwy_Num	Highway number.	Integer

Finally yet importantly, two index tables were created. The first index table was created to index the video files. The second index table was created to index entered data items for each highway. This index table is useful when editing, adding, or deleting data

¹⁵ This length was calculated using a custom-made Avenue script by the author that accounted for data projection.

Figure 4.6: Portion of the digital map after processing. Scale 1:500000.

Chapter 5:

CUSIS Synthesis

5.1. Introduction

CUSIS synthesis was conducted after the framework development, data collection, reformatting, and processing. Developing CUSIS within the *ArcView® GIS* environment was a complicated task, which required prior knowledge of computer programming conventions. The functionality of the interface controls, the windows components dimensions, the window resizing criteria, and such matters related to interface design, were acquired through personal experience and through consultation with thesis supervisors. On the other hand, the user side of the interface (i.e., user model) was tested and validated continuously through the help of graduate students in the *Department of Civil and Environmental Engineering* and thesis supervisors. Their feedback was considered in adjusting the CUSIS dialogs, responses, and output format.

Different integration methods of CUSIS within *ArcView® GIS* were studied. The objective was to find the optimum integration method. The available options were:

1. *Parent and child dialogs:* This integration method formats CUSIS as a hierarchy of dialogs, with the top most parent dialog (i.e., the home dialog) invoked automatically upon start-up, or by a custom button. Navigation through the system would be restricted to Parent/Child hierarchy, with the option of going to the home level. This method allows only certain combination of CUSIS-Modules, CUSIS-Applications, or CUSIS-Functions to be run simultaneously.

2. Menu driven: This method formats CUSIS as an extra menu. The menu would be placed just before the help menu, as the convention is that the last menu should be the help menu. This would intuitively convince the user, that *ArcView® GIS* has been expanded, and that CUSIS has added extra capabilities. The menu would be divided to five parts, where each part is separated by a menu-separator. Each portion of the menu would represent a CUSIS-Module, and each menu item would represent a CUSIS-Application. The different dialogs and procedures would be invoked by selecting a menu item. Because the menu is persistent, the user can simultaneously run many CUSIS-Applications (see Figure 5.1). This configuration is considered optimum, since it would not compromise the existing user interface of *ArcView® GIS*. In addition, users who are willing to perform regular *ArcView® GIS* functions will not be interrupted nor disturbed by the existence of CUSIS.
3. Separate buttons: This method would format CUSIS as many buttons. Each button would invoke a certain CUSIS-Application or CUSIS-Function. Using this method allows the user to invoke as many procedures as he/she can, however, it would be very confusing and not user friendly.

Figure 5.1: A screen shot showing the CUSIS menu within ArcView® GIS.

Five CUSIS-Applications were selected for inclusion in the final project. Two of these CUSIS-Applications are within the CAM—they are “Collision Database,” and “Collision Analysis.” The third selected CUSIS-Application was “Alignment Data,” which is a part of the CUSIS-Module HCM. The fourth CUSIS-Application was “Traffic Volume,” which is a part of the CUSIS-Module OCM. Finally yet importantly, the fifth CUSIS-Application was “Raw Video Clips,” which is a part of the VVM. Developing these five CUSIS-Applications involved creating eighty-three Avenue scripts, and thirteen *ArcView*® Dialogs. Some dialogs have more than one view, which in turn increased the number of user-viewable dialogs. The following sections describe the developed CUSIS-Applications and the engineering principles underlying them. Screenshots are provided for illustration.

5.2. “Collision Database” Application

The “Collision Database” is a CUSIS-Application that is a part of the CUSIS-Module named “Collision Analysis module.” When the user selects this menu item, a dialog appears (see Figure 5.2) showing the different CUSIS-Functions of this CUSIS-Application. Each CUSIS-Function is represented by a button. The button label describes the function and conventional help information is displayed in the status bar.

Figure 5.2: “Collision Database” main dialog.

5.2.1. “Plot a New Collision Table” Function

This CUSIS-Function allows the user to plot the exact location of individual collisions, for a newly added collision table. When the user clicks this button, the “Collision Database” main dialog, shown in the previous figure (Figure 5.2) will close, and then one of two scenarios will happen:

1. Scenario A: If there are collision tables that have not been plotted yet, this CUSIS-Function will prompt the user to choose the one to plot from them (see Figure 5.3). Then a set of *Avenue* scripts will locate the corresponding highway segments, and each individual collision will be exactly located along the corresponding highway segment, then a red dot will be placed in that location. Finally, a confirmation dialog appears after the collisions had been

plotted successfully, however, if the user selects the “Cancel” button, the operation will be aborted, and a warning message informs the user, that he/she has cancelled the operation.

Figure 5.3: Choice of highway to plot.

2. Scenario B: If there are no new collision tables, then an error screen appears informing the user that there were no new collision tables.

A new field is added to the attributes table of the “Data HWYs” theme. This new field contains the total number of collisions for every highway segment, which can be used in further analysis of the data. A problem in determining the exact location of a collision was identified because discrepancies exist in the reported lengths of highway segments, derived from the collision reports and GIS data layers. This discrepancy is due to combination factors:

- The underlying GIS map was digitised from *the Ministry of Transportation’s County Map Series* for Southern Ontario and from *the Provincial Map Series* produced by the Ministry of Natural Resources at a data capture scale of 1:100,000. The horizontal positional accuracy of the DCRB data is ± 100 metres.

- The base maps are two-dimensional maps, and precision in the road network database is determined by the skill and experience of the person digitising the road network.
- In the GIS software, roads and other features are referenced as two-dimensional features. In two dimensions, roads appear as if they were in plan view. Curves, tangent sections, and intersections appear as they would on an aerial photograph or map (i.e., an orthogonal view). Curves in a road obviously influence the measured distance between two nodes. The real world is three-dimensional topographic variations affect the distance along a roadway the same way horizontal curves do in two dimensions. The measurements taken by field personnel are obviously made in the real world and, therefore, accurately record distances as they are measured along both horizontal and vertical curves. These measurements reflect the geographic accuracy of a roadway and are the distances used to reference features in LHRS.

Therefore, GIS had to reflect the appropriate level of geographic accuracy, in other words, calibrating the distance to correspond to real world measurements. LHRS points were used as control points to calibrate the highway segment lengths in GIS. The following formula (equation 5.1) was used for calibration:

$$d_i = d_{AB} \times \frac{D_i}{D_{AB}} \quad (5.1)$$

Where: d_i = Calibrated distance for collision i

D_i = Measured GIS distance for collision i

d_{AB} = Calibrated distance between control points A and B

D_{AB} = Measured GIS distance between control points A and B

Through this process, each individual collision was exactly marked along the corresponding highway (see Figure 5.4 and Figure 5.5). Placement of the exact location of a collision was achieved by use of dynamic segmentation, which is a set of GIS Processes through which linear referenced data could be placed along a measured line or route system and spatial attributes to be derived from that location placement (Smith, Harkey, and Harris 2001, 30).

Figure 5.4: Collision distribution along highways 7, 15, 17, and 41 (MTO eastern region).

Scale 1:1,370,038.

Figure 5.5: Collision locations at intersection of highways 7 and 41. Scale 1:100,000.

5.2.2. “Query Collision Database” Function

“Query Collision Database” is a CUSIS-Function that is a part of the CUSIS-Application “Collision Database.” It allows the user to query the collision records, and display query results, or export these results to a *dBase*® database file. When the user clicks the “Query Collision Database” button on the “Collision Database” main dialog, the dialog closes and a new dialog titled “Collision Database -- Query” opens (see Figure 5.6). Through this dialog (i.e., “Collision Database -- Query”) the user chooses a highway

to query, and the *ComboBoxes*¹ are automatically loaded with all possible data values that the user can query in addition to the “All” option.

Collision Database -- Query			
Display collisions for Highway No. 7 based on the following criteria			
LHRs	All	Road Location	All
Year	All	Alignment	All
Month	All	Environment	All
Day of Week	All	Surface Conditions	All
Day of Month	All	Driver Action	All
Class	All	Driver Condition	All
No. of Vehicles	All	Vehicle Type	All
Initial Impact	All	Vehicle Maneuver	All
Light Condition	All		
<input type="button" value="New Set"/> <input type="button" value="Select From Set"/> <input type="button" value="Add to Set"/> <input type="button" value="Cancel"/>			

Figure 5.6: “Collision Database -- Query” dialog.

When the user selects criteria for a query (e.g., LHRs number, Year, Month, and so forth), the program seamlessly builds a SQL² select statement. The user can then select one of four buttons (i.e., options) to use. The first button, “New Set,” enables the user to build a new query and temporarily store the query results for further processing. The second button, “Select From Set,” enables the user to narrow the query criteria, and select from a previously selected data set. The third button, “Add to Set,” enables the user to expand the query criteria, and add the results to a previously selected data set. Finally, the fourth button, “Cancel,” closes the dialog and exits CUSIS. After running a query, the

¹ “A ComboBox control provides an interface for selecting an object from a list of related objects. A ComboBox consists of a set of rows, each of which is associated with a single object of any kind. One row is always presented to the user. This is the selected row. When the user wants to select another row, they click on a part of the combo box in order to popup a single-column, scrolling list that shows all rows in the combo box.” (ESRI 1996a)

² SQL stands for Structured Query Language. An industry-standard language for creating, updating and, querying relational database management systems.

results are displayed in a table (see Figure 5.7). The data records can be sorted in an ascending or descending order using any data field as the sort key. In addition, the query results can be exported to a dBaseIV file (“Export” button), or the user can return to the previous dialog by pressing the “Back” button. If the export option is chosen, another dialog opens (see Figure 5.8) prompting for the export file location.

Query Results										
Lhrs	Offset (Km)	Microfilm	Date	Time	Day	Class	Vehicle No.	Mainline	Initial Impact	L
14024	1.00	61011456	19961111	110	Mon	PDOnly	VehNo-01	Mainline	SnglVeh	D ▲
14026	0.00	60321966	19960226	145	Mon	PDOnly	VehNo-01	Mainline	SnglVeh	D
14026	0.80	61071488	19961125	1745	Mon	PDOnly	VehNo-01	Mainline	SnglVeh	D
14026	5.60	60750806	19960812	430	Mon	PDOnly	VehNo-01	Mainline	SnglVeh	D
14030	3.40	60490479	19960603	2135	Mon	PDOnly	VehNo-01	Mainline	SnglVeh	D
14035	0.60	60921929	19961028	630	Mon	Injury	VehNo-02	Mainline	RearEnd	D
14035	0.60	60921929	19961028	630	Mon	Injury	VehNo-01	Mainline	RearEnd	D
14035	1.70	60470069	19960527	1230	Mon	PDOnly	VehNo-04	Mainline	RearEnd	D
14035	1.70	60470069	19960527	1230	Mon	PDOnly	VehNo-03	Mainline	RearEnd	D
14035	1.70	60470069	19960527	1230	Mon	PDOnly	VehNo-01	Mainline	RearEnd	D
14035	1.70	60470069	19960527	1230	Mon	PDOnly	VehNo-02	Mainline	RearEnd	D
14035	2.40	60470067	19960527	1825	Mon	Injury	VehNo-03	Mainline	RearEnd	D
14035	2.40	60470067	19960527	1825	Mon	Injury	VehNo-02	Mainline	RearEnd	D
14035	2.40	60470067	19960527	1825	Mon	Injury	VehNo-01	Mainline	RearEnd	D ▼

Export

Back

Figure 5.7: “Query results” dialog.

Figure 5.8: Results export dialog.

5.2.3. “Add a New Collision Record” Function

“Add a New Collision Record” is a CUSIS-Function that is a part of the CUSIS-Application “Collision Database.” When the user clicks the “Add a New Collision Record” button on the “Collision Database” main dialog, the dialog closes and a new dialog titled “Collision Database -- Add” opens (see Figure 5.9). This dialog represents the data entry interface for collision database tables. All efforts were made to ensure that all possible and recognisable, data entry errors were eliminated. For example, if the user enters an offset greater than the section length, a wrong date, future date, time, or zero as the number of vehicles; the program displays an error message, explaining the reason for the error, and erases the wrong entries. The fields that have fixed set of values (as provided by MTO) are represented as *ComboBoxes*, to eliminate the possibility of the user entering a wrong code. In addition, the “Day” field is calculated automatically using the entered date. The user always has the option of clearing all entered data values by clicking “Reset” button. Furthermore, the “Continue” button is not enabled unless all fields have been filled.

Collision Database -- Add

Highway No.	7	Initial Impact	Angle
LHRS	14024	Light Condition	Dark
Offset (Km)		Road Location	AtIntSect
Microfilm No.		Alignment	CrvHil
Date (yyyymmdd)		Environment	Clear
Time (hhmm)		Surface Conditions	Dry
Day		Driver Action	Disobey Traf Cont
Class	Fatal	Driver Condition	Drinking
No. of Vehicles		Vehicle Type	Ambulance
Mainline	Mainline	Vehicle Manoeuvre	ChangeLane

Continue Reset Cancel

Figure 5.9: “Collision Database -- Add” dialog.

If all of the required data fields are entered correctly (i.e., without recognisable errors), and the user clicks the “Continue” button, the current dialog closes and a new set of dialogs open in turns (see Figure 5.10). Each dialog prompts the user to enter up to three events (i.e., series of events from the beginning of the collision until the vehicle had come to rest) for each vehicle. The first event field is compulsory, however, the second and third event fields are optional and the user can continue without entering any data. Then the program uses dynamic segmentation to plot a point in the exact location of the collision. Finally, a confirmation message will be displayed, such that the user will know that it was a successful operation.

Enter Vehicle Events

Vehicle No. 1

Event 1

Event 2 &--

Event 3 &--

OK Cancel

Figure 5.10: “Enter Vehicle Events” dialog.

5.2.4. “Delete a Collision Record” Function

“Delete a Collision Record” is a CUSIS-Function that is a part of the CUSIS-Application “Collision Database.” When the user clicks the “Delete a Collision Record” button on the “Collision Database” main dialog, the dialog closes and a new dialog titled “Delete a collision record” opens. The user has to select an entry from each *ComboBox* in the shown order (see Figure 5.11). Then the program prompts the user for a password to authorise this action (see Figure 5.12). The password is masked, which enhances the security of the system. Notice that this is the only CUSIS component that requires a password, as arbitrarily deleting a record would compromise data integrity.

Figure 5.11: “Delete a collision record” dialog.

Figure 5.12: “Password” prompt.

5.3. “Collision Analysis” Application

According to Persaud, Lyon, and Nguyen (1999, 7), it is important that collision analysis be efficient. Resources should not be wasted on sites that are incorrectly identified as unsafe, while unsafe sites are left untreated if they are not properly identified. Conventional techniques utilising collision counts, rates, or frequencies, often

in a statistical framework, have difficulties identifying sites because of the bias due to the *regression-to-the mean*³ phenomenon. The EB technique aims to smooth out the random fluctuation in collision data by specifying the safety of a site as an estimate of its long-term mean instead of its short-term count.

The “Collision Analysis” is a CUSIS-Application that is a part of the CUSIS-Module CAM. This CUSIS-Application analyses the collision history for a specific highway segment, highway, part of a highway network, or the entire highway network. It predicts collisions for a site, and then smoothes the prediction using the historic collision data by implementing an EB procedure. When the user selects this menu item, a dialog opens under the title “Collision Analysis Method”. There are two methods of analysis to select—that is intersection analysis, or highway segment analysis (see Figure 5.13). The first method (i.e., intersection analysis) analyses the collision data at intersections and interchanges. The second method (i.e., highway segment analysis) analyses the collision data along a highway, based upon segment-by-segment analysis. Through both methods, the user can set the scope of the analysis to a highway, a region in the highway network, or the entire highway network (see Figure 5.14).

Figure 5.13: “Collision Analysis Method” dialog.

³ “Regression to the mean is a technical term in probability and statistics. It means that, left to themselves, things tend to return to normal, whatever that is.” (Thornley 1997)

Figure 5.14: “Choose Scope of Analysis” dialog.

This CUSIS-Application is built upon an analysis method that was first introduced by Persaud, Lyon, and Nguyen (1999), and refined by Ontario Ministry of Transportation (Province of Ontario n.d.). This analysis method implements an Empirical Bayes procedure for ranking sites for safety investigation by potential of safety improvement. This method calculates a *potential of improvement index (POI)* for each site according to the following steps:

1. When the user selects the “Analyse” button on the “Choose Scope of Analysis” dialog (see Figure 5.14), this dialog is closed, and the collision numbers for each highway segment in the scope of the analysis are extracted. Each highway segment is divided into a number of sub-segments, such that the sub-segment length is as close as possible to one kilometre. Then the collision numbers and traffic volume data are extracted from the collision database, and the traffic volume data-table, respectively. This is done according to the recommendation of *Ontario Ministry of Transportation* (Province of Ontario n.d., 23), to avoid studying individual segments in isolation.
2. After extracting collision counts, another dialog titled “OP Factors” automatically opens (see Figure 5.15). This dialog prompts the user for the

factors, which will be used by the operational performance functions⁴ to estimate the operational performance for each highway within the scope of the analysis. These factors are derived by the Ontario Ministry of Transportation (Province of Ontario n.d., 23) for each class of highway, intersection, or interchange. The user has to press the “OK” button to enter the factors values, otherwise, if the “Cancel” button is pressed, or the dialog is closed, then any change to the default values will be ignored. MTO provides the OP factor values for each type of highway or intersection (Province of Ontario n.d.).

Figure 5.15: “OP Factors” dialog.

3. Using the traffic volume data, and the operational performance factors, the operational performance for sub-segments is estimated for each type of collisions using the following equations (equations 5.2 and 5.3):

$$OP = (\text{Section length in kilometres}) \times (a (AADT)^b) \quad (5.2)$$

$$OP = a (Mainline AADT)^b \quad (5.3)$$

⁴ “Operational Performance Functions (OPF’s) relate the operational performance of a highway entity (measured in collisions/year) to the traffic and geometric variables. The function gives the expected number of collisions (by type or severity level) for a highway entity based on the highway entity’s AADT [...]. The OPF is expressed in collisions/km/year for a highway segment or non-intersection element. It is expressed in collisions/year for intersections and interchanges.” (Province of Ontario n.d., 19)

4. To deal with the randomness in historical collision data, existing collision numbers are smoothed, for every type of collisions, using (equation 5.4) that is based on the Empirical Bayes theory (Province of Ontario n.d., 24), as follows:

$$m = \frac{\left(1 - \frac{k}{k + (n \times OP)}\right) \times count}{n} + \left(\frac{k}{k + (n \times OP)} \times OP\right) \quad (5.4)$$

where:

m - smoothed number of collisions

$count$ - number of collisions

OP - expected collisions according to operational performance function

n - number of years of collision data.

k - Operational performance function parameter

5. For a given highway site (i.e., highway segment, intersection, or interchange), the *potential for operational improvement* (POI) is defined as the difference between the smoothed number of collisions, and the expected operational performance. The POI for each highway site is estimated by collision type (i.e., fatal, injury, or property damage only), as illustrated by the following equation (equation 5.5). However, if the POI value for a highway site is

negative (i.e., this site is experiencing fewer collisions than expected) it is assigned a value of zero (Province of Ontario n.d., 25).

$$POI_i = m_i - OP_i; \text{ where } (i) \text{ is the collision type.} \quad (5.5)$$

6. A dialog titled “Severity Weightings” is automatically opened (see Figure 5.16), for the user to enter the weights (SW) of collision types as multiples of PDO collisions. The default values are 3.5 for injury collisions, and 140 for fatal collisions (Province of Ontario n.d., 25).

Figure 5.16: “Severity Weightings” dialog.

7. The POI index is calculated using the entered severity weightings, as a weighted average of the POI for each collision type (equation 5.6). Results are displayed in a table (see Figure 5.17), where they can be ranked from the highest to the lowest POI index using the sort tools provided. The highway site with the highest POI index is the site recommended for safety diagnostic activities. In addition, the user has the option to export the results as a dBaseIV file.

$$POI_{index} = (SW_{fatal} \times POI_{fatal}) + (SW_{injury} \times POI_{injury}) + POI_{PDO} \quad (5.6)$$

Collision Analysis Results

Analysis Type: HWY Segment

Scope: HWY 7

ID	HWY	From	To	Length (Km)	POI	POI per Km
1	7	14024+0.000	14024+1.235	1.235	0.0000	0.0000
2	7	14024+1.235	14024+2.470	1.235	0.2205	0.1786
3	7	14026+0.000	14026+1.066	1.066	0.7280	0.6832
4	7	14026+1.066	14026+2.131	1.066	0.2316	0.2173
5	7	14026+2.131	14026+3.197	1.066	0.0000	0.0000
6	7	14026+3.197	14026+4.263	1.066	0.0000	0.0000
7	7	14026+4.263	14026+5.329	1.066	0.0000	0.0000
8	7	14026+5.329	14026+6.394	1.066	0.0175	0.0164
9	7	14026+6.394	14026+7.460	1.066	0.0000	0.0000

Export POI per Stretch Cancel

Figure 5.17: Analysis Results dialog.

8. This method was extended to calculate the POI index for a stretch of the highway. Practically, the highway jurisdictions will investigate a stretch of the highway, rather than a one-kilometre section (Easa, Hassan, and Siczkar 1999). When the user presses the “POI by Stretch” button, an interim dialog opens for the user to enter the length of the required stretch. The default stretch length is ten kilometres. The user always has the option to go back to the original results and recalculate the POI index for another stretch length.

5.4. “Alignment Data” Application

“Alignment Data” is a CUSIS-Application that is a part of the CUSIS-Module “Highway Characteristics Module.” This CUSIS-Application provides the capability of drawing horizontal and vertical alignments through GIS. This was made possible by the

linking tabulated alignment data⁵ and spatial data. To invoke this CUSIS-Application, the user has to select the menu item labelled “Alignment Data.” A dialog, titled “Alignment Data – Choose Segment” opens and the user is prompted to use a tool that was specially developed by the author for picking up highway segments information (see Figure 5.18). This tool will fill in the highway number field, and the segment starting LHRs number field. If the user clicks on the map where there are no highway segments, an error message appears advising the user to try again. In addition, if the user chooses a highway segment that is missing necessary data, an error message appears advising the user of that error.

Figure 5.18: “Alignment Data - Choose Segment” dialog.

If the user successfully chooses a highway segment that has the necessary data, and presses the “Continue” button, the current dialog closes and another dialog, titled “Plot Options” opens (see Figure 5.19). Through this dialog, the user can choose to plot horizontal alignment, vertical alignment, or cross-section; however, the corresponding plot option is dimmed when there is no available data. If the chosen highway segment is located within two geographic townships, which is rare in the data used in this research, the user has the option to plot all of the section or to choose the portion passing by a

⁵ Refer to section “4.4 Data Reformatting.”

certain township. In addition, the user has the option to cancel the entire process, or to choose another highway section.

Figure 5.19: “Plot Options” dialog.

When the user presses the “Graph” button, on the “Plot Options” dialog (Figure 5.19), a new dialog, titled “XY Graphing⁶,” opens (see Figure 5.20). Through the “XY Graphing” dialog, the user can set the various properties of the physical graph. The available properties include; page width and height, margins, plot area, minimum and maximum values for the X-scale, minimum and maximum values for the Y-Scale, and the graph titles. The “Plot Options” dialog (Figure 5.19) remains open to facilitate further production of alignments, until the user presses the “Cancel” button. Plots are produced as layouts⁷, which could be printed directly from *ArcView® GIS*, or exported as an image file for further processing. Horizontal and vertical alignments are produced with the same horizontal scale to facilitate comparison and matching between alignments. Furthermore, horizontal curves are numbered in series from the beginning of the LHRS section, so the

⁶ XY Graphing is originally developed by Prof. Danny E. Patterson as a general-purpose graphing extension for ArcView® GIS. It was modified by the author to adopt the graphing requirements of CUSIS.

⁷ A Layout is an ArcView GIS document. “Layouts allow you to combine views, tables, charts and other images in a form which can be readily printed.” (Patterson 2002, 13)

user can easily cross-reference the plots to the tabulated alignment data (results for Highway 41 are shown in the next chapter).

XY Graphing (ver 1.1)					
Page Properties					
Width	11.0	Height	8.5		
X margin	2.0	Y margin	2.0		
X length	6.5	Y length	5		
[Calculate Graph Properties]					
Graph Properties					
X min	-10	X max	95	X inc	953
Y min	130	Y max	330	Y inc	20
X axis title:	Station				
Y axis title:	Elevation (m)				
Graph title:	Vertical:Grattan Township - HWY 41 - LH				
Info		Graph		Cancel	

Figure 5.20: “XY Graphing” dialog.

5.5. “Traffic Volume” Application

“Traffic Volume” is a CUSIS-Application that is a part of the *Operational Characteristics Module*. This CUSIS-Application is the querying interface for the traffic volume data. To invoke this CUSIS-Application, the user has to click the menu item labelled “Traffic Volume.” When this CUSIS-Application is invoked, a dialog is opened that has a tool for choosing the highway segment to query (see Figure 5.21). This tool behaves in the same manner as the “Location Tool” discussed earlier in section 5.4 (“Alignment Data” Application). The next step is to choose a year or average value across all years; this is done using the *ComboBox* on the same dialog. Then the corresponding AADT figure is displayed automatically. This CUSIS-Application can provide AADT data to other CUSIS modules or applications (e.g., providing AADT data to “Collision Analysis” application).

Figure 5.21: “Choose Location and Year” dialog.

5.6. “Raw Video Clips” Application

“Raw Video Clips” is a CUSIS-Application that is part of the “Video View” CUSIS-Module. This CUSIS-Application provides the user with the ability to view a video clip for a specific horizontal curve, which is useful when conducting a comprehensive safety evaluation for a highway segment (see Figure 5.22). The scope of this CUSIS-Application included highways 12, 31, 41, 43. The videos were shot from the driver’s eye height by mounting a digital video camera inside a passenger vehicle, and driving with the posted speed limit of the corresponding highway.

Figure 5.22: "Video View Module -- Raw Videos" dialog.

5.7. Building a Relational Database

To enhance the efficiency and performance of CUSIS, the data tables had to be linked together (see Figure 5.23). The linking method used within *ArcView® GIS* is the same as that of relational database management systems. Five one-to-many relationships were built between six data tables. The first relationship was built between the "Data HWYs" theme data table, with the "Segment" field as a primary key, and the "Collision Points" theme data table, with "Segment field" as a foreign key. The second through the fifth relationships were built between the "Collision Points" theme data table, with the field "Microfilm" as a primary key, and the four collision data tables (for the four highways being considered), with the "Microfilm" field as a foreign key. Hence, when the user manually selects or runs an operation that selects data from any of these tables,

the corresponding records will be automatically selected, for enhanced visual perception of the data.

Figure 5.23: Relationships between data tables.

Chapter 6:

CUSIS Implementation

6.1. Introduction

CUSIS was implemented to illustrate its analytical capabilities. Although not all CUSIS-Modules were developed, it provides the user with the necessary analytical tools to identify collision prone locations, and a fair contribution to the safety diagnosis efforts. Furthermore, this implementation would serve as a means, through which CUSIS could be functionally verified. Highway 41 was chosen for the purpose of this analysis. This highway runs South from its intersection with *Highway 2* (Geographic Township of Richmond, Lennox and Addington County), to North where it intersects *Renfrew County Road 19* (Geographic township of Pembroke, Renfrew County). The implementation process was done on four steps:

1. Collision data analysis and identifying collision prone locations;
2. querying the collision data base and the alignment data base to identify collision patterns and related conditions;
3. viewing videos of potential hazardous sites; and
4. formulating the findings into a report.

6.2. Collision Analysis

The CUSIS-Application “Collision Analysis” was run on Highway 41. It was assumed that the scope of the analysis would be a seven-kilometre stretch of this highway. The analysis results were sorted in a descending order using the POI index (see Figure 6.1). The section starting with LHRs 29580+9.917 and ending with LHRs

29590+5.045, had the highest total POI of 7.659. While the section starting with LHRs 29650+3.880 and ending with LHRs 29660+1.020 had the second highest total POI of 7.504. On the other hand, when the sections were ranked by the POI per kilometre, the order between those two sections was reversed with a difference of only 0.007. When the maps of these two sections were studied, it was found that the LHRs 29650 represented the intersection of Highways 41 and 132 (see Figure 6.2). It is evident from the literature that intersections (at large) compromise highway safety, therefore, the section starting with LHRs 29650+3.880 was considered for further analysis. The preceding and following segments were also included in the analysis to avoid studying this segment in isolation.

Collision Analysis Results

Analysis Type: HWY Segment (7.000 Km Stretch)

Scope: HWY 41

ID	HWY	From	To	Length (Km)	POI	POI Per Km
5	41	29580+9.917	29590+5.045	7.029	7.659	1.090
16	41	29650+3.880	29660+1.020	6.840	7.504	1.097
17	41	29660+1.020	29660+8.160	7.140	7.377	1.033
13	41	29630+4.875	29630+11.700	6.825	7.082	1.038
15	41	29640+7.000	29650+3.880	6.880	6.980	1.015
1	41	29570+0.000	29570+7.117	7.119	6.793	0.954
20	41	29690+7.156	29700+4.778	6.824	6.103	0.894
18	41	29660+8.160	29680+3.800	6.940	6.063	0.874
7	41	29600+0.992	29600+7.933	6.944	5.296	0.763
6	41	29590+5.045	29600+0.992	7.046	5.038	0.715
9	41	29610+2.970	29610+9.900	6.930	4.948	0.714
8	41	29600+7.933	29610+2.970	6.938	4.160	0.600
21	41	29700+4.778	29710+3.000	6.824	3.757	0.551
4	41	29580+2.975	29580+9.917	6.944	3.645	0.525
14	41	29640+0.000	29640+7.000	7.000	3.315	0.474

ExportPOI per StretchBack

Figure 6.1: Collision analysis results sorted by POI in a descending order.

Figure 6.2: Intersection of highways 41 and 132.

6.3. Querying the Databases

A query performed by the CUSIS-Function “Query Collision Database” for the initial impact type (Figure 6.3), resulted in 91 percent single vehicle collisions (Figure 6.4). Further refinement of the query results using CUSIS (Figure 6.5), concluded that 85 percent of these single vehicle collisions (which is almost 77 percent of total collisions) were due to hitting a wild animal, running off road, or skidding and sliding. Furthermore, 8 percent of the single vehicle collisions that correspond to “Ditch,” Cable guide rail,”

“Snow pile,” and “Pole-sign/park” must have first skidded or ran off road. As for the road conditions during these collisions, 52 percent of the running off road, and skidding/sliding collisions, happened at dry surface conditions, 21 percent in wet conditions, and 27 percent happened when the road surface was covered with slush (3%), loose snow (3%), pack snow (9%), or ice (12%). Finally, 52 percent of collisions caused by hitting a wild animal took place in dark conditions, 32 percent in daylight conditions, and 12 percent in dusk or dawn conditions.

Query Results										
Lhrs	Offset	Microfilm	Date	Time	Day	Class	Vehicle No.	Mainline	Initial Impact	Light
29660	8.30	91201715	19991220	1230	Mon	Injury	VehNo-01	Mainline	SnglVeh	Daylig
29660	6.50	91261330	19991206	1630	Mon	PDOOnly	VehNo-01	Mainline	SideSwipe	Daylig
29660	6.50	91261330	19991206	1630	Mon	PDOOnly	VehNo-02	Mainline	SideSwipe	Daylig
29660	6.50	91261330	19991206	1630	Mon	PDOOnly	VehNo-03	Mainline	SideSwipe	Daylig
29660	7.10	90740209	19990616	1100	Wed	PDOOnly	VehNo-01	Mainline	SnglVeh	Daylig
29660	6.00	90460264	19990305	1100	Fri	Injury	VehNo-01	Mainline	Turning	Daylig
29660	6.00	90460264	19990305	1100	Fri	Injury	VehNo-02	Mainline	Turning	Daylig
29660	6.00	80491569	19980525	1025	Mon	PDOOnly	VehNo-01	Mainline	Turning	Daylig
29660	6.00	80491569	19980525	1025	Mon	PDOOnly	VehNo-02	Mainline	Turning	Daylig
29660	2.30	80491362	19980522	630	Fri	PDOOnly	VehNo-01	Mainline	SnglVeh	Daylig
29660	2.30	80500686	19980522	2350	Fri	PDOOnly	VehNo-01	Mainline	SnglVeh	Dark
29660	0.60	80500580	19980515	1900	Fri	Injury	VehNo-01	Mainline	SnglVeh	Daylig
29660	0.20	80430224	19980424	100	Fri	PDOOnly	VehNo-01	Mainline	SnglVeh	Dark
29660	8.10	71041900	19971026	2320	Sun	PDOOnly	VehNo-01	Mainline	SnglVeh	Dark

Figure 6.3: Collision database query results for Highway 41.

Figure 6.4: Distribution of initial impact type.

Figure 6.5: Distribution of first event for single vehicle collisions.

Figure 6.6: Road surface conditions for running off road, and skidding/sliding collisions.

Figure 6.7: Light conditions for collisions due to hitting a wild animal.

Finally, the CUSIS-Application “Alignment Data” was used to generate horizontal and vertical alignments (see Figure 6.8). The mountainous terrain of these sections was clear and evident from the generated alignments. In addition, it was obvious that combining vertical and horizontal alignments at the time of design was not considered, which means that three-dimensional highway alignment was not considered, and the driver’s perception of curves was not studied.

Figure 6.8: Horizontal and vertical alignments for LHR 29650.

6.4. Viewing Highway Videos

Four horizontal curves were considered for video logging. Each curve was shot in both driving directions. The videos emphasised the findings reported earlier from the rest of the CUSIS-Applications and functions. Furthermore, it was clear that the highway surface was deteriorated which would compromise the side friction supplied by the road, which in turn would lead to running off road and sliding/skidding collisions. Finally, the mountainous terrain was clear and that heavy forests are very close to the highway premises, which in turn makes the appearance of a wild animal a surprise to the driver, which in turn makes him/her loose control (see Figure 6.9 and Figure 6.10). From this example, the potential benefits for the *Pavement Characteristics Application* are evident, as it would provide among other features the surface friction data that would lead to better conclusions.

Figure 6.9: A photo captured from one of the videos, showing the mountainous terrain and the close trees.

Figure 6.10: A photo captured from one of the videos, showing one of the sharp horizontal curves.

6.5. Suggested Countermeasures

The following countermeasures are suggested to avoid safety problems that may arise from the conditions and attributes pointed to by the analyses:

- Installing a wire fence to stop wild animals from going through the highway.
- Reconstructing the surface layer of the highway to increase its serviceability and friction properties.
- Installing warning signs for limited sight distance sections.
- Revising and modifying the highway alignment.

Chapter 7:

Summary, Conclusions, and Recommendations

7.1. Summary

The highway transportation system is a very complex and diverse system. One could be overwhelmed with the interoperable features of the highway system, hence, might lose track of the right attributes or factors to consider when making a decision. Highway jurisdictions, engineers, and professionals, are susceptible to overlooking the highway system, thus, they may make a wrong decision. Maintaining a safe transportation system is a strategic goal for all highway agencies. Introducing a safety improvement is one of many decisions the highway professional has to make. A powerful and comprehensive decision support tool, is needed by the highway professional to assist in making effective and timely decisions.

This thesis described research work conducted at Carleton University to develop a GIS-based safety information system. A comprehensive literature review of current and up-to-date safety analysis methods and techniques was presented. Based upon this comprehensive literature review, a set of goals were established that effectively addressed the user needs and requirements. These goals were used later as the corner stones of CUSIS, in order to avoid the pitfalls identified by other researchers.

A thorough study of GIS software framework development was conducted, to fully understand and acquire the essential knowledge to develop a robust framework for CUSIS. An object-oriented analysis was conducted to identify user needs, requirements, and system characteristics. An expandable framework was developed, that would

enhance the development of CUSIS. The developed framework was modular in nature, which means it could be modified with minimal efforts. In addition, a study was made to choose the candidate development environment. *ArcView® GIS* was chosen as the development environment for its object-oriented program model, flexibility, ease of customisation, and affordability for deployment.

CUSIS was developed and integrated as an extra menu in *ArcView® GIS*. Data acquired from MTO were reformatted and/or processed to suit CUSIS. Despite the fact that all of the data used within CUSIS were from MTO, CUSIS was developed as a generic scalable system. CUSIS is not data dependant, in terms of development; rather it is data-format dependant. This means that it would work with any data from any other jurisdiction, however, the data have to be formatted to CUSIS requirements. For the time limitations of the thesis, CUSIS development was limited to five CUSIS-Applications. The rest of CUSIS is to be developed later.

Finally, CUSIS was implemented on Highway 41, where it proved its uniqueness, proficiency, and ability to provide robust decision support. Results from the analysis were further refined, and counter measures were suggested (though suggesting countermeasures is out of the thesis scope). At the end, summary, conclusions, and recommendations for future work were presented.

7.2. Conclusions

A state-of-the-art safety information system is developed by integrating analytical techniques and GIS technologies. The developed system overcame the disadvantages of current safety analysis techniques and methods. The developed system is generic, easy to

use, user friendly, modular, and modifications could be performed with minimal effort. This system enables visual and analytical comprehension of collision-related data. It provides the user with robust and efficient decision-support. The following conclusions can be drawn:

- This thesis presents the development of *Carleton University Safety Information System (CUSIS)*, which is a GIS-based safety information system.
- CUSIS is a state-of-the-art safety information system, which is considered the first system to step foot in the trail of integrated highway safety analysis.
- CUSIS provides *all-in-one* safety analysis tool, where the discrepancy between the state-of-the-art and the state-of-practice was eliminated.
- CUSIS reveals its uniqueness by considering an Empirical Bayesian technique for collision analysis, which smoothes out the random fluctuation in collision data by specifying the safety of a site as an estimate of its long-term mean instead of its short-term count.
- CUSIS succeeded in bringing the diverse GIS tools and techniques into effect to the engineering world.
- CUSIS is the first safety information system that fuses safety-related data and utilise them in safety analysis.
- The object-oriented model used to develop CUSIS, simplifies problem handling, data management, as well as the capability of changing the system or expanding it with minimal effort.

- The user interface abides by the conventions of computer programming, and identifies, where applicable, most if not all of the identifiable data entry errors.
- The user interface considers logical conditions; data structure limitations, and input format requirements, when accepting data from the user.
- Development of CUSIS illustrates the well-established claim of power to GIS technologies.
- Results from CUSIS implementation proved that it conforms to its established goals, and it achieved its intended contributions to the highway safety science. Furthermore, a comprehensive collision analysis is illustrated, where results are easy to understand, hence, CUSIS proved its usefulness in the highway safety realm.
- A feasibility study is needed to verify the cost-effectiveness of CUSIS.
- Finally, another study is needed to investigate the organisational impact of implementing CUSIS in transportation jurisdictions.

7.3. Recommendations for Future Work

CUSIS has been developed to work in a standalone configuration. Adopting a client/server or a web-based configuration would increase its efficiency. The connectivity features offered by the web are incomparable with traditional ways of transferring data. Furthermore, a web-based configuration offers central administration, without compromising the deployment scope. The prevailing recommendation is CUSIS completion and deploying it in transportation jurisdictions. Besides adopting a web-based

configuration and completing the rest of the modules, CUSIS would benefit from adding extra modules, namely:

- A *Diagnostics and Remedy Module* that offers automated diagnosis of analysis results and that would suggest appropriate countermeasures.
- A *Data Check and Auto Format Module* to serve as a user interface for adding batch data to CUSIS. This module would check data format, and if possible suggests automatic corrections to the data format and coding.
- *IHSDM Link Module* that would link CUSIS to IHSDM, hence, benefiting from IHSDM analytical capabilities. This module would provide an online link to IHSDM, and an option to export the highway data in IHSDM input format.

Apart from adding new modules to CUSIS, further enhancement and refinement of the existing modules is also recommended. It is recommended to expand the *Alignment Data Application* to include a method of converting vertical profile to a series of tangents and parabolic curves (Easa, Hassan, and Abdul Karim 1998). This would enhance the analysis of vertical profiles, hence, leading to better analysis results. In addition, it is recommended to expand the analytical capabilities of CUSIS to include all types of rural highways. Considering the Highway Safety Manual within CUSIS core would greatly enhance its analytical capabilities. In addition, developing a variant of CUSIS to analyse urban roads is recommended.

Apart from enhancing and expanding CUSIS, using GPS to record collision data is heavily recommended. Collisions are linear, not point, events by nature. Recording the

full path for each involved vehicle from point where the collision started to the point where it ended, would provide the safety professional with a full understanding of individual collisions, hence representing each collision with a line (or series of lines) is recommended. This recommendation implies adding a *Collision Import and Simulation Module* for acquiring GPS data and simulating the collision scenario.

References

- Al-Masaeid, Hashem R. 1997. Performance of Safety Evaluation Methods. *Journal of Transportation Engineering* 123(5): 364-369.
- Anderson, Ingrid B., Karin M. Bauer, Douglas W. Harwood, and Kay Fitzpatrick. 1999. Relationship to Safety of Geometric Design Consistency Measures for Rural Two-Lane Highways. *Transportation Research Record* 1658: 43-51.
- Australia. Department of Defence. Defence Science and Technology Organization. Centre for Cognitive Work and Safety Analysis. 2003. *Analytic Approaches*. [cited 7 Jun. 2003] <<http://www.dsto.defence.gov.au/ssl/cwsa/cwa.html>>.
- Boutilier, Bob. 2002. Interview by author. Ontario Ministry of Transportation. Geometric Design Section. Kingston, Ontario. Jun. 2002.
- Brown, Bruce, and Karsten Baass. 1997. Seasonal Variation in Frequencies and Rates of Highway Accidents as Function of Severity. *Transportation Research Record* 1581: 59-65.
- Carreker, Lauren Elise, and William Bachman. 2000. Geographic Information System Procedures to Improve Speed and Accuracy in Locating Crashes. *Transportation Research Record* 1719: 215-218.
- Chowdhury, Mashrur A., Nicholas J. Garber, and Duan Li. 2000. Multiobjective Methodology for Highway Safety Resource Allocation. *Journal of Infrastructure Systems* 6(4): 138-144.
- Coad, Peter, and Edward Yourdon. 1991. *Object-Oriented Analysis*. 2nd ed. Yourdon Press Computing Ser. Englewood Cliffs, NJ: Prentice Hall.

- Council, Forrest M., and J. Richard Stewart. 2000. Safety Effects of the Conversion of Rural Two-Lane Roadways to Four-Lane Roadways. *ITE Journal* 70(4): 37-41.
- Council, Forrest M., Yusuf M. Mohamedshah, and J. Richard Stewart. 1997. Effects of Airbags on Severity Indexes for Roadside Objects. *Transportation Research Record* 1581: 66-71.
- Crossroads Software. 2003. *Traffic Collision Database*. [cited 7 Jun. 2003] <<http://www.crossroadssoftware.com/products/traffic/tcdframes.htm>>.
- Dana, Peter H. 2003. Geodetic Datum Overview. In *The Geographer's Craft Project*. The University of Colorado at Boulder, Department of Geography. [updated 11 Feb. 2003; cited 17 Jun. 2003] <http://www.colorado.edu/geography/gcraft/notes/datum/datum_f.html>.
- Davis, Gary A., and Shimin Yang. 2001. Bayesian Identification of High-Risk Intersections for Older Drivers via Gibbs Sampling. *Transportation Research Record* 1746: 84-89.
- Dictionary of Computing*. 1997. Ed. Valerie Illingworth. 4th ed. London, UK: Oxford University Press.
- Digital Cartographic Reference Base of Ontario. Ver. 3.0. Geomatics Office. Ontario: Ministry of Transportation.
- Dissanayake, Sunanda, J. John Lu, and Xuehao Chu. 1999. Simplified Approach to Forecast Highway Crash Rates of Selected Special Population Subsets. *Transportation Research Record* 1665: 44-50.

- Easa, Said M., Yasser Hassan, and Dennis Siczkar. 1999. An Approach for Auditing Highway Sections for Safety Improvements. *Canadian Journal of Civil Engineering* 26: 724-735.
- Easa, Said M., Yasser Hassan, and Z. Abdul Karim. 1998. Establishing Highway Vertical Alignment Using Profile Field Data. *ITE Journal* 68(8): 81-86.
- Edwards, Dave. 2002. Interview by author. Province of Ontario. Ministry of Transportation. Traffic Section. Kingston, Ontario. Jun. 2002.
- Elvik, Rune, and Anne Borger Mysen. 1999. Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries. *Transportation Research Record* 1665: 133-140.
- English, John. 1997. *Ada 95: The Craft of Object-Oriented Programming*. London, UK: Prentice Hall Europe.
- Environmental Systems Research Institute, Inc. (ESRI) 1996a. *Avenue: Customization and Application Development for Arcview GIS*. Redlands, CA: ESRI Press.
- . 1996b. *Using the ArcView Network Analyst*. Redlands, CA: ESRI press.
- . 1997. *Arcview Dialog Designer*. Redlands, CA: ESRI Press.
- . 2002. Geography Matters. *ESRI® Whitepaper*. Redlands, CA: ESRI Press.
- . 2003. Spatial Data Standards and GIS Interoperability. *ESRI® Whitepaper*. Redlands, CA: ESRI Press.
- Forrester, Jay W. 1991. System Dynamics and the Lessons of 35 Years. In *The Systemic Basis of Policy Making in the 1990s*. Ed. Kenyon B. De Greene. N.p.
- Gebers, Michael A. 1998. Exploratory Multivariable Analyses of California Driver Record Accident Rates. *Transportation Research Record* 1635: 72-80.

- Graettinger, Andrew J., Timothy W. Rushing, and John McFadden. 2001. Evaluation of Inexpensive Global Positioning System Units to Improve Crash Location Data. *Transportation Research Record* 1746: 94-101.
- Griffith, Michael S. 2000. Safety Evaluation of Rolled-In Continuous Shoulder Rumble Strips Installed on Freeways. *ITE Journal* 70(6): 37-42.
- Hall, James P., Tschangho John Kim, and Michael I. Darter. 2000. Cost-Benefit Analysis of Geographic Information System Implementation: Illinois Department of Transportation. *Transportation Research Record* 1719: 219-226.
- Harkey, David L., and Michael S. Griffith. 1999. GIS-Based Crash Referencing and Analysis System. *ITE Journal* 69(4): 9-12.
- Harlow, Charles, and Yu Wang. 2001. Automated Accident Detection System. *Transportation Research Record* 1746: 90-93.
- Harwood, Douglas W. 1995. Relationships Between Operational and Safety Considerations in Geometric Design Improvements. *Transportation Research Record* 1512: 1-6.
- Harwood, D. W., F. M. Council, E. Hauer, W. E. Hughes, and A. Vogt. 2000. *Prediction of the Expected Safety Performance of Rural Two-Lane Highways*. FHWA-RD-99-207. McLean, VA: Federal Highway Administration.
- Hassan, Yasser. 2002. *Geometric Design: CIVE5308 Course Notes*. Carleton University.
- Hauer, Ezra. 1996a. Statistical Test of Difference Between Expected Accident Frequencies. *Transportation Research Record* 1542: 24-29.
- . 1996b. Detection of Safety Deterioration in a Series of Accident Counts. *Transportation Research Record* 1542: 38-43.

- . 1996c. Identification of Sites with Promise. *Transportation Research Record* 1542: 54-60.
- . 1997. *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Tarrytown, N.Y: Pergamon.
- . 2001. Computing and Interpreting Accident Rates for Vehicle Types or Driver Groups. *Transportation Research Record* 1746: 69-73.
- Heylighen, F. 1997. Occam's Razor. In *Principia Cybernetica Web*. Ed. F. Heylighen, C. Joslyn and V. Turchin. Brussels, Belg.: Principia Cybernetica. [updated 7 Jul. 1997; cited 5 Jun. 2003] <<http://pespmc1.vub.ac.be/OCCAMRAZ.html>>.
- Hopkins, John B. 2002. *Realistic Assessment of New Transportation Technologies and System Concepts: Even If You Manage to Build It, They May Not Come*. Applications of advanced technologies in transportation: Proc. of the seventh international conference on applications of advanced technology in transportation, August 2002, Cambridge, Massachusetts. Ed. Kelvin C. P. Wang, Samer Madanat, Shashi Nambisan, and Gary Spring. Reston, Virginia: American Society of Civil Engineers.
- JMW Engineering, Inc. (JMW). 2003. *AIMS Capabilities*. [cited 7 Jun. 2003] <http://www.jmwengineering.com/aims_capability.htm>.
- Jones, Andrew P., and Stig H. Jørgensen. 2003. The use of multilevel models for the prediction of road accident outcomes. *Accident Analysis and Prevention* 35(1): 59-69.

- Kanellaidis, George. 1999. Aspects of Road Safety Audits. *Journal of Transportation Engineering* 125(6): 481-486.
- Khan, Sarosh, Ramalingam Shanmugam, and Brian Hoeschen. 1999. Injury, Fatal, and Property Damage Accident Models for Highway Corridors. *Transportation Research Record* 1665: 84-92.
- Kilov, Halim, and William Harvey, eds. 1996. *Object-Oriented Behavioral Specifications*. Boston, MA: Kluwer Academic Publishers.
- Krammes, Raymond A., and Carl Hayden. 2003. Making Two-Lane Roads Safer. *Public Roads* 66(4): 37 pars. [cited 6 Jun. 2003] <<http://www.tfhrc.gov/pubrds/03jan/04.htm>>.
- Lamm, Ruediger, Artur K. Guenther, and Elias M. Choueiri. 1995. Safety Module for Highway Geometric Design. *Transportation Research Record* 1512: 7-15.
- Lamm, Ruediger, Basil Psarianos, Elias M. Choueiri, and George Soilemezoglou. 1995. *A Practical Safety Approach to Highway Geometric Design International Case Studies: Germany, Greece, Lebanon, and The United States*. International Symposium on Highway Geometric Design Practices, August - September 1995, Boston, Massachusetts.
- Lang, Laura. 1999. *Transportation GIS*. Redlands, CA: Environmental Systems Research Institute, Inc.
- Leore, Robert C. 2000. Canadian Highway Information System. Presentation to Transport Canada. Ottawa, ON. 15 Apr. 2000.
- . 2003. Interview by author. Transport Canada. Ottawa, Ontario. 4 Jun. 2003.

- Liang, Wei Lien, Michael Kyte, Fred Kitchener, and Patrick Shannon. 1998. Effect of Environmental Factors on Driver Speed. *Transportation Research Record* 1635: 155-161.
- Lo, C. P. and Albert K. W. Yeung. 2002. *Concepts and Techniques of Geographic Information Systems*. Prentice Hall Ser. in Geographic Information Science, ed. Keith C. Clarke. Upper Saddle River, NJ: Prentice Hall.
- Lowry, Richard. [2003]. *Concepts and Applications of Inferential Statistics*. Poughkeepsie, NY: Vassar College [cited 29 May 2003] <<http://faculty.vassar.edu/lowry/webtext.html>>.
- Mak, King K. 1995. Safety Effects of Roadway Design Decisions—Roadside. *Transportation Research Record* 1512: 16-21.
- Martiz, J. S. and T. Lwin. 1989. *Empirical Bayes Methods*. New York, NY: Chapman and Hall.
- McGinnis, Richard G., Matthew J. Davis, and Eric A. Hathaway. 2001. Longitudinal Analysis of Fatal Run-Off-Road Crashes, 1975 to 1997. *Transportation Research Record* 1746: 47-58.
- Mehmood, Arif, Frank Saccomanno, and Bruce Hellinga. 2001. Simulation of Road Crashes by Use of Systems Dynamics. *Transportation Research Record* 1746: 37-46.
- Mendoza, Alberto, Emilio F. Mayoral, José L. Vicente, and Francisco L. Quintero. 2001. Geographic Information System-Based Accident Data Management for Mexican Federal Roads. *Transportation Research Record* 1746: 74-83.

- Mensah, Abraham, and Ezra Hauer. 1998. Two Problems of Averaging Arising in the Estimation of the Relationship Between Accidents and Traffic Flow. *Transportation Research Record* 1635: 37-43.
- Mercier, Cletus R., Mack C. Shelley II, Julie B. Rimkus, and Joyce M. Mercier. 1997. Age and Gender as Predictors of Injury Severity in Head-on Highway Vehicular Collisions. *Transportation Research Record* 1581: 37-46.
- Miaou, Shaw-Pin, An Lu, and Harry S. Lum. 1996. Pitfalls of Using R^2 to Evaluate Goodness of Fit of Accident Prediction Models. *Transportation Research Record* 1542: 6-13.
- Miller, John S. 1997. Using Technology To Help Overcome Institutional Obstacles To Improved Crash Records Processing. *Transportation Research Record* 1581: 10-17.
- . 2000. Geographic Information Systems: Unique Analytic Capabilities for the Traffic Safety Community. *Transportation Research Record* 1734: 21-28.
- Miller, Ted, David T. Levy, Rebecca S. Spicer, and Diane C. Lestina. 1998. Allocating the Costs of Motor Vehicle Crashes Between Vehicle Types. *Transportation Research Record* 1635: 81-87.
- Moffatt, Mike. 2003. Glossary of Economics Terms. In *About.Com Guide to Economics*. [cited 6 Jul. 2003] <<http://economics.about.com/library/glossary/blglossary.htm>>.
- Mohamedshah, Yusif M., Li Wan Chen, Forrest M. Council. 2000. Association of Selected Intersection Factors with Red-Light-Running Crashes. *ITE Journal* 70(7): 37-42.

- Morrall, J. F., and R. J. Talarico. 1994. Side Friction Demanded and Margins of Safety on Horizontal Curves. *Transportation Research Record* 1435: 145-152.
- Mountain, Linda, Bachir Fawaz, and David Jarrett. 1996. Accident Prediction Models for Roads With Minor Junctions. *Accident Analysis and Prevention* 28(6): 695-707.
- Mouskos, Kyriacos C., Wu Sun, Steven I. Chien, Arthur Eisdorfer, and Tao Qu. 1999. Effect of Midblock Access Points on Traffic Accidents on State Highways in New Jersey. *Transportation Research Record* 1665: 75-83.
- National Research Council (NRC[USA]). Transportation Research Board. Committee for a study for a future strategic highway research program. 2001. Strategic Highway Research: Saving Lives, Reducing Congestion, Improving Quality of Life. *Transportation Research Board Special Report* 260. Washington, DC: National Academy Press.
- Nierstrasz, Oscar. 1989. A Survey of Object-Oriented Concepts. In *Object-Oriented Concepts, Databases, and Applications*. Ed. Won Kim and Frederick H. Lochovsky. New York, NY: ACM Press.
- Norin, Hans, and Irene Isaksson-Hellman. 1995. Injury Potential Prediction of a Safety Design Feature: A Theoretical Method Based on Simulations and Traffic Accident Data. *Safety Science* 19(1): 45-56.
- O'Looney, John. 1997. *Beyond Maps: GIS and Decision Making in Local Government*. Washington, DC: International City/Council Management Association.
- Panchanathan, Sriram, and Ardesbir Faghri. 1995. Knowledge-Based Geographic Information System for Safety Analysis at Rail-Highway Crossings. *Transportation Research Record* 1497: 91-100.

- Paniati, Jeffery F., and Forrest M. Council. 1997. The Highway Safety Information System: Transforming Data Into Knowledge. *Public Roads* 60(3): 4-9.
- Patterson, D. E. 2002. Geographic Information Systems: A Primer. N.p.
- . 2003. Working With ArcView GIS: Beyond The Basics. N.p.
- Pd' Programming, Inc.(Pd). 2003. Intersection Magic Home Page. [cited 7 Jun. 2003] <<http://www.pdprog.com/im/>>.
- Persaud, Bhagwant N. 2001. *Statistical Methods in Highway Safety Analysis*. NCHRP Synthesis of Highway Practice 295. Washington, DC: National Academy Press.
- Persaud, Bhagwant N., Craig Lyon, and Thu Nguyen. 1999. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. *Transportation Research Record* 1665: 7-12.
- Persaud, Bhagwant N., and Leszek Dzbik. 1993. Accident Prediction Models for Freeways. *Transportation Research Record* 1401: 55-60.
- Persaud, Bhagwant N., and Thu Nguyen. 1998. Disaggregate Safety Performance Models for Signalized Intersections on Ontario Provincial Roads. *Transportation Research Record* 1635: 113-120.
- Persaud, Bhagwant N., R. Retting, P. Garder, and D. Lord. 2001. Observational Before-After Study of the Safety Effect of U.S. Roundabout Conversions Using the Empirical Bayes Method. Presented at the 80th Annual Meeting of the Transportation Research Board. Washington, DC.
- Pierce, B. K., and J. G. Kinader. 1999. Incorporating Geographic Correlation When Sampling a Transportation Network. *Transportation Research Record* 1665: 13-21.

- Pietrucha, Martin T., Timothy R. Pieples, and Philip M. Garvey. 2000. Evaluation of Pennsylvania Road Safety Audit Pilot Program. *Transportation Research Record* 1734: 12-20.
- Polus, A., J. Craus, M. Livneh, and S. Katznelson. 1999. Analysis of Flow, Safety and Warrants for Paved Shoulders on Two-Lane Rural Highways. *Road and Transport Research* 8(1): 42-55.
- Province of Ontario. Ontario Ministry of Transportation (MTO). Road Safety Program Office. 2003. *Ontario Road Safety Annual Report 2001*. Toronto, Ontario: Publications Ontario.
- . ---. Traffic Office. Traffic Safety Management Section. n.d. *The Science of Highway Safety*. N.p.
- . ---. ---. Traffic Information Management Section. [2000?] Notes on Traffic Volume Information System (TVIS) History Listing (Up To and Including 2000). N.p.
- Quiroga, Cesar A., and Darcy Bullock. 1996. Geographic Database for Traffic Operations Data. *Journal of Transportation Engineering* 122(3): 226-234.
- Razavi, Amir H. 1999. *Arcview GIS/Avenue Developer's Guide*. 3rd ed. Santa Fe, NM: Onword Press.
- Razavi, Amir H., and Valerie Warwick. 2000. *Arcview GIS/Avenue Programmer's Reference : Class Hierarchy Quick Reference and 100+ Scripts*. 3rd ed. Albany, NY: OnWord Press.
- Riggs, James L., David D. Bedworth, Sabah U. Randhawa, and Ata M. Khan. 1997. *Engineering Economics*. 2nd Canadian ed. Toronto, ON: McGraw-Hill Ryerson Limited.

- Riley, Sarah J. 2002. *Development of a Public Transit Information System Using GIS and ITS Technologies*. Master of Applied Science Thesis. Carleton University.
- Roget International Thesaurus Indexed Edition*. 2001. Ed. Barbara Ann Kipfer; Robert L. Chapman, consulting editor. 6th ed. New York: HarperCollins.
- Saccomanno, F. F., K. C. Chong, and S. A. Nassar. 1997. Geographic Information System Platform for Road Accident Risk Modeling. *Transportation Research Record* 1581: 18-26.
- Saccomanno, F. F., S. A. Nassar, and J. H. Shortreed. 1996. Reliability of Statistical Road Accident Injury Severity Models. *Transportation Research Record* 1542: 14-23.
- Sadek, Salah, Mounia Bedran, and Isam Kaysi. 1999. GIS Platform for Multicriteria Evaluation of Route Alignments. *Journal of Transportation Engineering* 125(2): 144-151.
- Sayed, Tarek. 1997. The Highway Safety Expert System: Diagnosing Accident Prone Locations. *Civil Engineering Systems* 14(4): 251-267.
- Sayed, Tarek, and Walid Abdelwahab. 1997. "Using Accident Correctability to Identify Accident-Prone Locations." *Journal of Transportation Engineering* 123(2): 107-113.
- Sayed, Tarek, Walid Abdelwahab, and Frank Navin. 1995. Identifying Accident-Prone Locations Using Fuzzy Pattern Recognition. *Journal of Transportation Engineering* 121(4): 352-358.
- Sinnott, Richard O. 1998. Frameworks: The Future of Formal Software Development? *Computer Standards and Interfaces* 19: 375-385.

- Smith, Richard C., David L. Harkey, and Bobby Harris. 2001. *Implementation of GIS-Based highway safety analyses: Bridging the gap*. FHWA-RD-01-039. McLean, VA: Federal Highway Administration.
- Souleyrette, Reginald R., and Daniel J. Gieseman. 1999. Development of an Automated Crash-Location System for Iowa. Presented at *AASHTO's GIS-T*. San Diego, CA.
- Souleyrette, Reginald R., Tim Strauss, Brad Estochen, and Michael Pawlovich. 1998. *GIS-Based Accident Location and Analysis System (GIS-ALAS)--Project Report: Phase I*. N.p.
- Spolsky, Joel. 2001. *User Interface Design for Programmers*. 1st ed. Berkeley, CA: APress.
- Spring, Gary S. 2002. Knowledge-Based Geographic Information System (GIS) for The Identification, Analysis, and Correction of Hazardous Highway Locations. In *Applications of advanced technologies in transportation: Proc. of the seventh international conference on applications of advanced technology in transportation, August 2002*. Cambridge, Massachusetts. Ed. Kelvin C. P. Wang, Samer Madanat, Shashi Nambisan, and Gary Spring. Reston, Virginia: American Society of Civil Engineers, 2002.
- Stewart, J. Richard. 1996. Application of Classification and Regression Tree Methods in Roadway Safety Studies. *Transportation Research Record* 1542: 1-5.
- Stokes, Robert W., and Madaniyo I. Mutabazi. 1996. Rate-Quality Control Method of Identifying Hazardous Road Locations. *Transportation Research Record* 1542: 44-48.

- Sun Microsystems, Inc. (SUN). 2003. Object-Oriented Programming Concepts. In *The Java Tutorial: A Practical Guide to Programmers*. [updated 19 May 2003; cited 10 Jun. 2003] <<http://java.sun.com/docs/books/tutorial/java/concepts/index.html>>.
- Tabachnick, Barbara G., and Linda S. Fidell. 2001. *Using Multivariate Statistics*. 4th ed. Boston, MA: Allyn and Bacon.
- Tarko, Andrzej P., Kumares C. Sinha, and Omer Farooq. 1996. Methodology for Identifying Highway Safety Problem Areas. *Transportation Research Record* 1542: 49-53.
- The Canadian Oxford Dictionary*. 2001. Don Mills, ON: Oxford University Press.
- The Chicago Manual of Style*. 14th ed. 1993. Chicago, IL: Chicago University Press.
- Thielman, Carol Y., and Michael S. Griffith. 1999. Overview of Three Expert Systems for Crash Data Collection. *Transportation Research Record* 1665: 147-157.
- Thomas, Fridtjof. 2003. Statistical Approach to Road Segmentation. *Journal of Transportation Engineering*. 129(3): 300-308.
- Thornley, David H. 1997. *Regression to the Mean*. [cited 4 Jul. 2003] <<http://www.visi.com/~thornley/david/philosophy/thinking/mean.html>>.
- Transport Canada. 2000. *Canadian Motor Vehicle Collision Statistics*. Canada: Minister of public works and government services.
- . Canadian Council of Motor Transport Administrators. 2001. *Road safety vision 2010*. Canada: Minister of public works and government services.
- Tsoukalas, Lefteri H., and Robert E. Uhrig. 1997. *Fuzzy and Neural Approaches in Engineering*. fwd. Lotfi A. Zadeh. New York, NY: John Wiley and Sons, Inc.

- United States of America (USA). Department of Transportation. Federal Highway Administration (FHWA). 1996. *Crash Referencing and Analysis System – User Manual*. FHWA-RD-00-146. N.p.
- , ---, ---. 2003a. *Interactive Highway Safety Design Model: Making Safety a Priority in Roadway Design*. [cited 6 Jun. 2003] <<http://www.tfhrc.gov/safety/ihsdm/ihsdm.htm>>.
- , ---, ---. Turner-Fairbank Highway Research Center (TFHRC). 2003b. *Highway Safety Information System*. [cited 5 Jun. 2003] <<http://www.hsisinfo.org/hsis.cfm?num=2&page=1>>.
- , ---, ---, ---. 2003c. SafetyAnalyst Tools. [cited 2 Jul. 2003] <<http://www.safetyanalyst.org/tools.htm>>.
- . United States Geological Survey. 2003d. *Geographic Information systems*. [cited 6 Jun. 2003] <<http://www.usgs.gov/research/gis/title.html>>.
- Vogt, Andrew, and Joe Bared. 1998. Accident Models for Two-Lane Rural Segments and Intersections. *Transportation Research Record* 1635: 18-29.
- Washington, Simon. 1999. Conducting Statistical Tests of Hypotheses: Five Common Misconceptions Found in Transportation Research. *Transportation Research Record* 1665: 1-6.
- Zegeer, Charles V. 1982. Highway Accident Analysis Systems. *NCHRP Synthesis of Highway Practice* 91. Washington, DC: National Academy Press.
- Zegeer, Charles V., and Forrest M. Council. 1995. Safety Relationships Associated with Cross-Sectional Roadway Elements. *Transportation Research Record* 1512: 29-36.

- Zegeer, Charles V., Herman F. Huang, J. Richard Stewart, and Carolyn Williams. 1998a. Investigation of National Highway System Roadways in the Highway Safety Information System States. *Transportation Research Record* 1635: 1-9.
- Zegeer, Charles V., Richard Stewart, Forrest Council, and Timothy R. Neuman. 1994. Accident Relationships of Roadway Width on Low-Volume Roads. *Transportation Research Record* 1445: 160-168.
- Zegeer, Charles V., Herman F. Huang, J. Richard Stewart, Ron Pfefer, and Jun Wang. 1998b. Effects of Tow Away Reporting Threshold on Crash Analysis Results. *Transportation Research Record* 1635: 49-57.
- Zhang, Dongsong, Lina Zhou, and Jay F. Nunamaker Jr. 2002. A Knowledge Management Framework for The Support of Decision Making in Humanitarian Assistance/Disaster Relief. *Knowledge and Information Systems* 4: 370-385.
- Zhang, Haiyi, Barbro Back, and WanLei Zhou. 2001. Expert Systems - The Design And Implementation of a Traffic Accident Analysis System. *Lecture Notes in Computer Science* 2070: 476-481.
- Zhang, Jianping, and Junming Yang. 1997. Instance-Based Learning for Highway Accident Frequency Prediction. *Microcomputers in Civil Engineering* 12(4): 287-294.
- Zhaou, Min, and Virginia P. Sisiopiku. 1997. Relationship Between Volume-To-Capacity Ratios and Accident Rates. *Transportation Research Record* 1581: 47-52.