Relações entre a operação de tráfego e segurança viária

**Programa de Pós-graduação em
Engenharia de Transportes da EESC-USP**

**Exemplar definitivo (corrigido). O exemplar
original está disponível na CPG da EESC-USP**

São Carlos, 31/01/2019

RESOLUÇÃO CoPGr N ° 6018, DE 13 DE OUTUBRO DE 2011, artigo 5º

Tese apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do grau de Doutor em Ciências, Programa de Pós-Graduação em Engenharia de Transportes.

Área de Concentração: Planejamento e Operação de Sistemas de Transportes

Orientador: Prof. Antonio Clóvis Pinto Ferraz

São Carlos
Novembro
2018

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Andrade, Gustavo Riente.

a553r Relações entre a operação de tráfego e segurança viária / Gustavo Riente de Andrade; orientador Antonio Clóvis Pinto Ferraz. São Carlos, 2018.

Tese (Doutorado) - Programa de Pós-Graduação em Engenharia Transportes e Área de Concentração em Planejamento e Operação de Sistemas de Transportes-- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2018.

1. Acidentes. 2. Operação. 3. Vias arteriais. 4. Rodovias de pista dupla 5. Modelagem 6. Severidade. I. Título.

Relationship between traffic operations and road safety

**Programa de Pós-graduação em
Engenharia de Transportes da EESC-USP**

Exemplar definitivo (corrigido). O exemplar original está disponível na CPG da EESC-USP

São Carlos, 31/01/2019

RESOLUÇÃO CoPG N ° 6018, DE 13 DE OUTUBRO DE 2011, artigo 5º

A Dissertation presented to the graduate school of engineering of the University of São Paulo in fulfillment of the requirement for the degree of Doctor in Sciences.

Advisor: Prof. Titular Antonio Clóvis Pinto Ferraz

São Carlos
Novembro
2018

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Andrade, Gustavo Riente.

a553r Relationship between traffic operations and road safety / Gustavo Riente de Andrade; adviser Antonio Clóvis Pinto Ferraz. São Carlos, 2018.

Tese (Doutorado) - Programa de Pós-Graduação em Engenharia Transportes e Área de Concentração em Planejamento e Operação de Sistemas de Transportes-- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2017.

1. Crashes. 2. Operations. 3. Urban arterials. 4. Expressways. 5. Modeling. 6. Crash severity.
I. Título.

FOLHA DE JULGAMENTO

Candidato: Engenheiro **GUSTAVO RIENTE DE ANDRADE**.

Título da tese: "Relações entre a operação de tráfego e segurança viária"

Data da defesa: 18/12/2018.

Comissão Julgadora:

Resultado:

Prof. Titular **Antonio Clovis Pinto Ferraz**
(Orientador)
(Escola de Engenharia de São Carlos/EESC)

Aprovado

Profa. Dra. **Ageliki Elefteriadou**
(University of Flórida)

Aprovado

Prof. Dr. **William Miller Sampson**
(University of Flórida)

Aprovado

Prof. Dr. **Sivaramakrishnan Srinivasan**
(University of Flórida)

Aprovado

Profa. Dra. **Ruth Lorraine Steiner**
(University of Flórida)

Aprovado

Prof. Titular **Jose Reynaldo Anselmo Setti**
(Escola de Engenharia de São Carlos/EESC)

Aprovado

Coordenadora do Programa de Pós-Graduação em Engenharia de Transportes:

Profa. Associada **Ana Paula Camargo Larocca**

Presidente da Comissão de Pós-Graduação:

Prof. Associado **Luís Fernando Costa Alberto**

DEDICATION

To my family and to those whose ultimate sacrifice paved the way for safer roads

ACKNOWLEDGMENTS

I would like to express my gratitude to the many people who supported me and made this dissertation possible.

My special thanks to my advisors, Dr. Lily Elefteriadou and Dr. Coca Ferraz, for believing in this project from the beginning and welcoming me to their research groups.

I would also like to extend these acknowledgements to all other professors who I had the opportunity to collaborate with and contributed in my growth as a researcher and professional. Special thanks to Dr. José Setti, Dr. Cira Pitombo and Dr. André Cunha at USP, and to Dr. Siva Srinivasan, Dr. Yafeng Yin, Dr. Washburn and Dr. Ruth Steiner at UF.

My further appreciation goes to the Mc Trans Center, prof. Bill Sampson and Vishal Khanapure, for the financial support thought these years and for being a valuable lab where research and practice joined together. To the USDOT UTC Program under STRIDE Center, that funded part of this research. Special thanks to Ines Aviles-Spadoni.

I am also thankful to the São Paulo state transport agency ARTESP, toll highway association ABCR-SP and affiliated highway concessionaries, which kindly provided the data for this research. Thanks to the financial support from the CNPq, Brazil.

I would also like to thank Tectran Group, which gave me the opportunity to experience real world problems and formulate the research questions that motivated me to pursue a path to academia.

My acknowledgements go to everybody whose support was crucial the life as a graduate student. Special thanks to Heloisa Belo and Beth Ortega at USP and, at UF, to the wonderful staff at Weil 365 and UFTI – Mrs. Debra, Blair, Doretha, Keri and Jennifer. My thanks to supportive people from the International Center, my international adviser Mrs. Ethel Porras. Special thanks to Mrs. Debra Anderson, who kindly helped me navigate through the life in this new country.

A section of this acknowledgements is rightfully dedicated to all of those who have worked hard to build the bridge between University of São Paulo and University of Florida, in the form of a novel PhD Dual Degree Agreement. Special thanks to Dr. Ana Paula Larocca and Cesar Derisso at

USP, and to Nancy McIlrath at UF, who demonstrated the good will and determination to go much beyond their daily duties to make this possible.

Thanks to the many friends I made who not only helped but made this journey a fun ride: Magaly Romão, Fernando Hirose, Fernando Piva, Gaurav Sultania, Kiarash Fariborzi, Fabio and Renata Sasahara, Ash Omidvar, Marilo Martin-Gasulla, Yinan Zheng and Zhibin Chen.

My profound gratitude to my parents Silvestre and Regina, my sister Carolina and family members, who have always incentivized me and inspired to go higher and higher.

Finally, my loving thanks to my wife Karla and my son Rafael, whose tender souls and sheer determinations overcame every distance to support me in this endeavor.

RESUMO

ANDRADE, G. R. *Relações entre a operação de tráfego e segurança viária*. 97p. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos. 2018.

Desde antes da publicação do Highway Safety Manual, vários pesquisadores indicam a necessidade de se incorporar aspectos de operação de tráfego à análise de segurança viária.

A primeira parte deste trabalho desenvolveu e implementou em uma ferramenta computacional existente um método de otimização de tempos semafóricos que considera medidas de desempenho de operação, segurança e emissões simultaneamente. Uma análise de sensibilidade foi realizada para produzir conhecimento sobre os efeitos práticos e a ordem de relevância de 20 variáveis de entrada principais. O desempenho da programação semafórica em termos de redução do atrasos geralmente coincide com redução das emissões, embora às vezes às custas da segurança.

A segunda parte deste trabalho investigou a relação entre a densidade horária de tráfego e as taxas de acidentes em autoestradas e rodovias de pista dupla brasileiras com características diversas, com base em um banco de dados contendo mais de 20.000 registros de acidentes e mais de 35 milhões de observações de tráfego. As curvas resultantes para rodovias urbanas seguem um formato em U, com valores mínimos associados aos níveis de serviço B a C, enquanto que as relações para as rodovias rurais são contínuas e crescentes, sugerindo que rodovias rurais de baixo volume são mais seguras do que as de maior volume. A análise de outros fatores revelou que condições noturnas, segmentos de entrelaçamento e rodovias de pista dupla convencionais urbanas estariam relacionadas a maiores taxas de acidentes.

A terceira parte deste projeto amplia a análise para modelagem de severidade dos acidentes, usando um modelo de escolha discreta ordenado. A estrutura que melhor se adequa a esse banco de dados levou ao desenvolvimento de dois modelos diferentes: acidentes com um veículo e acidentes com múltiplos veículos, já que os fatores que explicam a severidade dos acidentes variam muito entre esses modelos. Por exemplo, defensas e barreiras se mostraram efetivas para a redução da severidade de acidentes com um veículo, para as quais a saída de pista é o tipo de acidente mais grave. O amplo banco de dados usado neste estudo também permitiu uma investigação da influência das condições de tráfego na severidade do acidente, em comparação com todos os outros

fatores. Os resultados sugeriram que a severidade de acidentes de múltiplos veículos está negativamente relacionada com a densidade de tráfego, enquanto colisões com um único veículo estão mais relacionadas à velocidade.

As descobertas deste trabalho têm implicações nas decisões sobre políticas e projetos de transportes, e a equação produzida pode ser incorporada à análise de confiabilidade do gerenciamento ativo do tráfego (ATM) e do Highway Capacity Manual.

Palavras-chave: Acidentes; operação; vias arteriais; rodovias de pista dupla; modelagem; severidade.

ABSTRACT

ANDRADE, G. R. *Relationship between traffic operations and road safety*. 97p. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos. 2018.

Since before the release of the Highway Safety Manual research has been indicating the need to incorporate mobility and control aspects to road safety analysis.

The first part of this work developed and implement in an existing computational engine a signal timing optimization method that considers mobility, safety, and emissions measures simultaneously. A sensitivity analysis was conducted to provide insight on the practical effects and order of relevance of 20 key input variables. Mobility improvement performance usually coincides with emissions improvements, but sometimes at the expense of safety.

The second part of this work investigated the relationship between hourly traffic density and crash rates on Brazilian expressways with different characteristics, based on a database containing over 20,000 crashes and more than 35 million traffic volume observations and. The resulting curves for urban expressways follow a U shape, with minimum values associated with LOS B to C, while the relationships for rural expressways were found to be continuously increasing, suggesting that low volume rural roads are safer than the higher volume ones. The analysis of other influencing factors revealed that nighttime conditions, weaving segments and urban multilane highways could be related to higher crash rates.

The third part of the project extends the analysis to crash severity modeling, using an ordered response choice model. The framework that better fit this database led to the development of two different models: single-vehicle crashes (SV) and multiple-vehicle crashes (MV), since the factors that explain the severity of crashes varies widely between these models. For instance, guardrails and barriers proved to effectively reduce severity for SV crashes, for which run-offs are the most severe crash type. The unique database used in this study also allowed for an investigation of the influence of prevailing traffic conditions on crash severity, while still controlling for all other factors. The results suggested that multiple-vehicle crash severity is negatively related with traffic density, while single-vehicle crashes are more closely related to speed.

The findings of this work have implications to policy and design decisions, and the produced equation could be incorporated to active traffic management (ATM) and HCM reliability analysis.

Keywords: Crashes; operations; urban arterials; expressways; modeling; crash severity.

ABBREVIATIONS

AADT - Annual average daily traffic

AASHTO - American Association of State Highway and Transportation Officials

AASHTO - American Association of State Highway and Transportation Officials

ATM - Active traffic management

EPA - Environmental Protection Agency

FEM - Factorial Effect Method

FHWA- Federal Highway Administration

FI - Fatal and injury crashes

GA - Genetic algorithm

HCM - Highway Capacity Manual

HCS - Highway Capacity Software

HSM - Highway Safety Manual.

LOS - Level of service

MNL - Multinomial Logit Model

MV - Multiple-vehicle crashes

MVMT - Millions of vehicle-miles travelled

NGSIM - Next Generation Simulation

NMA - National Motorists Association

OL - Ordered Logit Model

OP - Ordered Probit Model

P&B - Pedestrians and bicycles

PDO - Property-damage only crashes

RTOR - Right-turn on red

SV - Single-vehicle crashes

TRB - Transportation Research Board

v/c - Volume-to-capacity ratio

VMT - Vehicle-miles travelled

TABLE LIST

Table 2.1 - Selected Safety Model Details	25
Table 2.2 - Database size and characteristics of previous crash prediction work.	30
Table 2.3 - Database size and characteristics of previous crash severity work.....	35
Table 3.1 - Variables included in the sensitivity analysis and variation range.	43
Table 3.2 – Random trajectory generator.	47
Table 3.3 – Kruskal-Wallis test for all variables and performance measures.	52
Table 3.4 – Results for different optimization schemes, per intersection.	54
Table 4.1 – Frequency of main road characteristics, number of crashes and exposure	59
Table 4.2 – Discrete variables included in the database.....	62
Table 4.3 – Continuous variables included in the database.....	63
Table 5.1 – Influencing variables 2-Way ANOVA statistics.	70
Table 5.2 – Model coefficients and ANOVA statistics.	73
Table 6.1 – Results and significant variables for each model.	82
Table 6.2 – Probabilities estimated for each non-continuous variable.....	83
Table 6.3 – Results and significant variables – Alternative MV and P&B Models.	88

FIGURE LIST

Figure 2.1 - Existing density-crash rate models and LOS thresholds.	29
Figure 3.1 - Configuration, demand and initial signal timing on testing arterial.	40
Figure 3.2 - Factorial effects for each input parameter and performance measure.	49
Figure 4.1 – Studied freeways, multilane highways and main urban areas.....	57
Figure 4.2 – Distribution of crashes by class and severity.	60
Figure 5.1 – Relationship between traffic density and crash rate, by severity.	66
Figure 5.2 – Relationships between traffic density and crash rate, by class.	67
Figure 5.3 – Distribution of single-vehicle crashes for LOS A-E vs LOS F.....	68
Figure 5.4 – Relationships between density and crash rate, by influencing variable.....	71
Figure 5.5 – Analysis of standardized residuals.	74
Figure 5.6 – Application of the proposed model.	75
Figure 5.7 – Comparison of results from the proposed equation and the HSM.	76
Figure 6.1 – Comparison of results from the proposed equation and the HSM.	84
Figure 6.2 – Relationships between traffic density and crash severity distribution.	86
Figure 6.3 – Variability of the relationship between speed and SV crash severity.	87
Figure 6.4 – Relationship between density differences and MV crash severity.	87

TABLE OF CONTENTS

1	INTRODUCTION	21
2	LITERATURE REVIEW	23
2.1	Crash Prediction at Signalized Intersections.....	23
2.2	Crash Prediction on Expressways.....	25
2.2.1	Safety-Operations Relationship.....	26
2.2.2	Safety-Operations Modelling	27
2.2.3	Summary.....	28
2.3	Crash Severity Distribution on Expressways.....	31
2.3.1	Studies on modeling crash severity using discrete choice models	31
2.3.2	Summary.....	33
3	EFFECTS OF MOBILITY, SAFETY AND EMISSIONS ON SIGNAL TIMING OPTIMIZATION	36
3.1	Methodology	37
3.1.1	Signal Timing Optimization	37
3.1.2	Emissions Model	38
3.2	Test Scenarios	39
3.2.1	Base Scenario	40
3.2.2	Alternative Scenarios.....	41
3.3	Sensitivity Analysis	43
3.3.1	Application	46
3.3.2	Results	47
3.3.3	Discussion on the Trade-off between Measures.....	50
3.3.4	Ranking of Influencing Variables by Measure.....	51
3.3.5	Impact of Weight Schemes on Signal Timing.....	52

3.4	Final remarks.....	55
4	CRASH AND TRAFFIC DATABASE FOR EXPRESSWAYS.....	57
5	RELATIONSHIP BETWEEN CRASH RATE AND OPERATIONS ON EXPRESSWAYS.....	64
5.1	Relationship between Density and Crash Rate.....	65
5.2	Modelling the Relationship Between Density and Crash Rate	68
5.2.1	Influencing Factors	68
5.2.2	Proposed Model	72
5.3	Comparison Between the Proposed Model and the HSM.....	76
5.4	Discussion	77
5.5	Final Remarks.....	77
6	ESTIMATION OF CRASH SEVERITY ON EXPRESSWAYS.....	79
6.1	Methodology Overview.....	80
6.2	Results	81
6.3	Marginal Effects	83
6.3.1	Crash Type	84
6.3.2	Time and Environment.....	85
6.3.3	Geometry and Design.....	85
6.3.4	Traffic Characteristics.....	86
6.4	Alternative Model.....	88
6.5	Final remarks	89
7	CONCLUSIONS AND RECOMMENDATIONS.....	90
8	REFERENCES.....	93

1 INTRODUCTION

The Highway Safety Manual – HSM (AASHTO, 2011) has been the most cited reference for road safety studies. The core of its methodology is based on Safety Performance Functions, which relate the annual average daily traffic to the predicted number of crashes (Daniel and Maine, 2011). Similar to other existing models, the HSM models were developed using aggregate traffic measures and are more suitable for planning level analysis.

Since before the release of the HSM, some authors have been indicating the need to incorporate operational characteristics into road safety analysis making it possible to account for the effects of seasonal, daily or hourly flow and speed fluctuations on crash rates (Lord, Manar and Vizioli, 2005). The development of such relationships can enable practitioners to account for both operations and safety aspects in designing and operating highway facilities. These relationships seem to be affected by hourly traffic parameters, and may influence a variety of policy assessments, ranging from Active Traffic Management strategies to capacity increase decisions.

In this context, the goal of this research is to model the relationship between crash rates and different operating conditions for two different components of a road network: corridors with signalized intersections and expressway (uninterrupted) segments.

Chapter 2 – Literature Review – describes previous work on crash prediction models for signalized intersections and expressways, summarizing the main conclusions that are relevant to this research.

Chapter 3 – Effects of Mobility, Safety and Emissions on Signal Timing Optimization – describes the development, software implementation and testing of a method that accounts not only for mobility, but also safety and emissions, to optimize signal timings along corridors.

Chapter 4 – Crash and Traffic Database for Expressway Segments detail the unique large and varied database that was used in the two subsequent chapters of this dissertation.

Chapter 5 – Relationship Between Crash Rate and Operations on Expressways describes the development of crash prediction models that relate crash rates to hourly traffic density and other roadway and environmental conditions, based on a large and diverse database.

Chapter 6 – Estimation of Crash Severity on Expressways further develops this research, presenting crash severity distribution models that are compatible with the crash prediction models described in Chapter 4, and incorporate the effect of a wide range of characteristics to the severity distribution functions.

Chapter 7 – Conclusion closes this dissertation, highlighting the contributions of this research project.

2 LITERATURE REVIEW

2.1 CRASH PREDICTION AT SIGNALIZED INTERSECTIONS

A large amount of work has been conducted to investigate safety at signalized intersections, usually using Poisson or Negative Binomial distributions to describe the relationships between geometry and crashes. Since the 1990's, research started to incorporate signal timing control features to develop methods that are able to predict crashes as a function of both geometry and operational aspects (Poch and Mannering, 1996; Mitra et al., 2007; Lambert, 1992; Chin and Quddus, 2003; Wang and Abdel-Aty, 2006; Agbelie and Roshandeh, 2015). These include cycle length, phasing, control types (pre-timed versus adaptive), coordination, protected versus permitted left-turns and yellow and all-red intervals. Increasing the number of phases was found to increase the number of crashes in general, but providing a protected left turn on the arterial could improve safety in some instances (Poch and Mannering, 1996, Chin and Quddus, 2003, Agbelie and Roshandeh, 2015), while increasing rear-end crashes when used on minor streets (Wang and Abdel-Aty, 2006). Coordination was found to improve both mobility and safety conditions (Chin and Quddus, 2007). Enhanced display types have also been related to safety improvements. Although not used in the USA, studies on traffic signal countdown timers (Spigolon, 2010; Islam et al., 2017) have been showing that this configuration is able to lead to safer responses by drivers by reducing the "dilemma zone".

Since the 2000's, microscopic traffic simulation has been employed to assess the interactions of safety and operation at intersections, using conflict-based surrogate measures to estimate crashes (Archer 2004; Ozbay et al., 2008; Gettman et al., 2008; Essa and Sayed, 2015; Shahdah et al., 2015).

The Highway Safety Manual – HSM (AASHTO, 2010) presents comprehensive models to evaluate safety on rural and urban intersections, using multiple explanatory variables and regression to predict crashes. Although comprehensive in nature, the HSM methodology includes few signal control parameters. The work conducted by Turner for the New Zealand transport authority (Turner

et al., 2012) moved one step ahead by combining operational and geometry parameters to estimate crashes by type. In the resulting equations, a base local scaling factor is multiplied by a series of variables that modify the number of estimated crashes as a function of demand, operations, and geometry characteristics:

$$A_j = b_0 x_1^{b_1} x_2^{b_2} \dots x_n^{b_n} \quad 2.1$$

where A is the predicted rate for type j crashes in a 5-year period; bi are model coefficients; xi are the independent variables of the model; and b0 is a local adjustment factor. This model was selected for this study. The median of the b0 defaults calculated by Turner et al. (2012) for different cities was used in this study.

Table 2.1 shows all which variables and their respective coefficients that are used by the safety model to calculate the number of predicted crashes per intersection approach and crash type. For example, the number of rear-end crashes at a small intersection approach with no exclusive right turn, no split-phasing, a cycle lane and a nearby bus bay would be calculated as:

$$A_{rear-end, small} = b_0 \left(\frac{through}{traffic} \right)^{.447} \times \left(\frac{left\ turn}{Storage\ Length} \right)^{-.259} \times \left(\frac{lost}{time} \right)^{-3.424} \times 0.706 \times 1.309 \quad 2.2$$

Note that the multiplicative structure of Equation 2.1 allows for the future inclusion of additional factors, obtained from other studies. Potential variables could be HSM Crash Modification Factors (AASHTO, 2001), the number of phases (Chin and Quddus, 2007), maximum greens (Agbelie and Roshandeh, 2015) or the presence of countdown timers (Spigolon, 2010; Islam et al., 2017).

Table 2.1 - Selected Safety Model Details

Factors x_i	Right angle	Left-turn against	Rear-end			Loss-of-control	Other
			small	medium	large		
Through traffic	$x^{0.311}$		$x^{0.447}$	$x^{.496}$	$x^{0.985}$	$x^{0.541}$	$x^{0.262}$
Conflicting crossing traffic	$x^{0.362}$						
Left-turning traffic		$x^{0.155}$					
Degree of saturation		$x^{0.397}$				$x^{0.447}$	
Total number of lanes	$e^{(.356x)}$			$e^{(.243x)}$	$e^{(.459x)}$	$e^{(.144x)}$	
Number of through lanes		$e^{(0.352x)}$					
Total width							$x^{0.027}$
Left-turn storage length		$(1+x)^{-0.124}$	$(1+x)^{-0.259}$		$(1+x)^{-1.142}$		
Presence of shared-turns	1.19						1.26
Shared left-turn		0.72					
Exclusive right-turn lane			1.585				
Intersection depth (m)	$x^{0.602}$						
High speed (> 35 mph)				1.449	0.985	1.57	1.98
Cycle time (sec)	$x^{0.037}$	$x^{-0.683}$				$x^{-0.704}$	$x^{0.354}$
All-red (sec)	$x^{-0.636}$						
Free right turn on red				1.442	1.227	1.17	1.16
Lost time (sec)			$x^{-3.424}$	$x^{-0.209}$	$x^{-1.739}$		
Split-phasing	0.69		5.256	1.57	0.95	2.47	1.21
Full left-turn protection		0.71					
Coordinated	1.31						0.71
Advanced detector	2.06						0.44
Mast arm signal display	0.74						
Median island	0.67	1.22					
Merge on intersection exit						1.47	0.65
Cycle facilities		1.35	0.706	0.753	1.257		
Approach bus bay			1.309	0.908		1.6	1.27
Upstream parking						0.58	0.7
CBD area				0.9	0.819		1.83
Residential area						0.75	

2.2 CRASH PREDICTION ON EXPRESSWAYS

The literature review first describes studies that sought to examine relationships between road and operational characteristics on crash quantities or rates, normally by means of statistical analysis. Those studies contributed to the body of knowledge in the field by indicating the best methods to conduct such investigations and by suggesting candidate variables that could be added in a model.

Following such recommendations, the second subsection presents the evolution of studies that propose user-ready models that relate operational variables to crashes. Finally, the summary contrasts the main results and recommendations from the work described herein.

2.2.1 SAFETY-OPERATIONS RELATIONSHIP

Before the 2000's, the few authors that investigated the relationship between operations and crashes faced the lack of information at the time of the crash, resulting in the use of averages that may mask the effect of prevailing flow, speed and other contributing variables to crash risk (Mensah and Hauer, 1998). From the early 2000's, a handful of studies were able to show with statistical confidence that prevailing traffic and environmental characteristics influence crash occurrence.

Martin (2002) found that crash rates on French freeways were higher on weekends, when traffic is low and trucks are restricted. Severity was found to be greater when hourly traffic is lower and increasing the number of lanes was related to fewer crashes. However, by using relationships between the total prevailing traffic flow and the number of crashes, traffic speeds and densities were not the same within the samples segments with different number of lanes. According to the authors, functional forms that include the prevailing density or v/c offer a better characterization of the crash process (Lord et al., 2005).

A study that used data from California freeways (Golob and Recker, 2003) showed a connection between congested flow and increased crash risk. Fatal and injury crashes were grouped in one category labeled (FI), while the remaining incidents were grouped into a category labeled Property-Damage Only (PDO). The authors found that FI crashes were influenced more by the prevailing volume than by the prevailing speed. Collisions were related to temporal variability in speed. Likewise, by modeling the relationship between v/c and crash rates on freeways using the Poisson model, Daniel and Maina (2011) showed that geometry characteristics that affect capacity also affect safety.

Arona et al. (2015) went further, by investigating the possible effect of daylight, average traffic speed, occupancy, average headways and combinations of speed, relative speed and gaps in

establishing relationships between prevailing traffic conditions and crash rates on French urban freeways. The estimated performance functions for single vehicle crashes were found to be affected by speed, while multiple vehicle crashes are more closely related to occupancy, which is directly related to traffic density.

2.2.2 SAFETY-OPERATIONS MODELLING

The first studies to model the relationship between traffic flows and crash rates date from the 80's (Frantzeskakis and Iordanis, 1981; Ceder, 1982). The fitted total accident rate-hourly flow function followed the U-shaped configuration.

Kononov et al. (2011) proposed a sigmoidal function to model the crash rates in urban freeways in Colorado and California, which predicts slight increases in the number of crashes for low traffic flows and a sharp rise with increasing traffic. U-shaped functions were also used to describe the relationship between v/c and crash frequency on a freeway segment, a tunnel and a toll plaza in South Korea (Chang et al., 2000), as well as an urban freeway segment in Detroit (Zhou and Sisiopiku, 2014). In these studies, the influence of ramps was not accounted for and was incorporated into the error term of the calibrated models.

Studies conducted more recently used models based on traffic density for four urban regions of the USA (Harwood et al., 2013; Potts et al., 2014). Lower crash rates occur for moderate values of prevailing traffic density. The resulting models describing the relationship between crash rates and v/c or traffic density are U-shaped functions. As in earlier studies (Frantzeskakis and Iordanis, 1981; Ceder, 1982), this shape results from of a convex downward and a convex upward curve for single- and multi-vehicle accidents, respectively.

However, none of these studies propose models that can incorporate the statistically relevant factors to a single equation. Also, no study has investigated the influence of the weavings, merges and diverges to such relationships. Finally, there isn't to the present any effort to examine these relationships in the Brazilian context.

2.2.3 SUMMARY

Studies that investigate the relationship between safety and operations suggest that congestion has an impact on crash rates. Several studies indicate that crash rates also increase under low volume conditions, resulting in U-Shaped functions describing the relationship between crash rates and v/c or traffic density.

Regarding severity, the literature recommends grouping fatal and injury crashes (FI), the remaining incidents being Property-Damage Only (PDO) (Golob and Recker, 2003).

With respect to the best mobility measure to relate to crash rates, the literature indicates that the v/c or traffic density (Chang et al., 2000; Daniel and Maina, 2011; Lord et al., 2005) is preferable over flow. The use of traffic density has the additional advantage of better handling undersaturated and oversaturated flows (Martin, 2002).

Regarding best practices in collecting traffic data for studies to correlate crash occurrence to prevailing traffic operational conditions, the use of 5-min intervals was found to be ideal by a previous study (Pande et al., 2005). Also, combining observations from all lanes was better than the use of lane-by-lane data, because this not only captures across-lane variation in speed or volume but also allows the use of a larger data set for analysis, since data from the specific crash-lane is often missing.

The only two USA practice-ready models relating traffic density in pc/mile/lane to crash rates in crashes per million of vehicle-miles travelled (MVMT) are shown in Figure 2.1. The graph also shows the boundaries for the levels of service (LOS) for freeways and multilane highways, according to the Highway Capacity Manual (HCM) 6th Edition (TRB, 2016). Note that the order of magnitude of both curves is similar and that the lowest crash rates are associated with levels of service B to C.

Figure 2.1 - Existing density-crash rate models and LOS thresholds.

Table 2.2 provides an overview of the most relevant studies on the relationship between traffic operations and crash rates, listing both the dimensions of the databases that were used in each study and the variables considered. The last column shows the corresponding values for this study, as a point of comparison. Note that, while most studies used the relationship between measured flows and speeds to calculate the traffic density, Arona et al. (2015) obtained densities by measuring occupancies.

Most of these studies relied on a limited amount of data and/or include a limited number of roadway and environmental characteristics. Some studies rely on data from a single facility, which provides valuable insight but is insufficient to generate broader conclusions. As a result, in spite of the existent research on the effects on safety of many of the variables shown in Table 2.2, no study has incorporated statistically relevant factors into a single equation in the format presented in Figure 2.1. Also, past work didn't distinguish weaves, merges and diverges from basic segments, so that the possible effect of such design alternatives to safety wasn't assessed. Finally, most models focused on urban freeways, and possible differences on such relationships for rural freeways remains to be investigated.

Table 2.2 - Database size and characteristics of previous crash prediction work.

		Martin (2002)	Golob and Recker (2003)	Lord, et al. (2005)	Daniel and Maine (2011)	Potts et al. (2014)	Zhou and Sisiopiku (2014)	Arona et al. (2015)	This Study
Location		France	Southern California	Canada	New Jersey	5 US Large Cities	Detroit Area	Marseille France	São Paulo Brazil
Facility	Multilane Highways								X
	Freeway Segments	X	X	X	X	X	X	X	X
	Ramps								X
	Weaves								X
	Traffic Stations	92	NA	9	987	1226	36	208	187
Crash Data	Count	11,172	1,200	1,296	NA	42,740	7,731	292	21,778
	Severity	X	X	X		X	X	X	X
	Num of Vehicles	X		X			X		X
	Type		X	X			X		X
Traffic Interval	Count	2,900,000	1,200		NA	97,834,433	3,983,401	18,220,800	36,830,195
	Interval	6 min	30 sec	5 min	1 hour	15 min	15 min	6 min	5-15 min
Method		Negative Binomial	PCA NLCCAA	Poisson	Negative Binomial / Poisson	Regression	Regression	Logistic Regression	Regression
Traffic	Volume		X	X	X	X	X		X
	Truck Percentage				X	X	X		X
	Free Flow Speed				X		X		X
	Avg. speed		X	X		X	X	X	X
Environment	Lighting/Time	X	X	X			X	X	X
	Weekend	X		X		X	X		X
	Work Zones					X	X		X
	Rural Areas				X				X
	Urban Areas	X	X	X	X	X	X	X	X
Geometry	Number of Lanes	X		X		X	X	X	X
	Lane/Shoulder Width				X		X	X	
	Design Speed						X		X
	Toll Station						X		X
	Total Ramp Density				X		X		X
	Guardrails / Barriers						X		X

PCA Principal Component Analysis

NLCCAA Nonlinear Canonical Correlation Analysis

2.3 CRASH SEVERITY DISTRIBUTION ON EXPRESSWAYS

Several approaches have been used to investigate crash severity and the factors that affect it. While several nonparametric models have been used to model the discrete nature of the severity output, such as the boosted regression tree and other tree-based models (Lee and Li, 2015), most studies have used discrete response models to address this topic. These are discussed in the next subsections.

2.3.1 STUDIES ON MODELING CRASH SEVERITY USING DISCRETE CHOICE MODELS

The ordered probit (OP) and logit models (OL), already common in econometric applications, have been used in severity analysis over several years (O'Donnell and Connor, 1996; Kockelman and Kweon, 2002; Abdel-Aty, 2003; Yamamoto and Shankar, 2003), using data from roadway segments, intersections and toll plazas. In general, crashes were found to be more serious for older drivers, males, speeding drivers, and passengers without seat-belt or involved in angle crashes. For the case of road segments, sharp curves and poor lighting were found to contribute to more severe crashes. The authors of these studies suggested that the reason more serious crashes occur in rural areas may be due to the higher speeds, though speed or other prevailing traffic conditions were unknown.

Other studies have used different types of discrete choice models, such as the mixed logit model (MLM), and the concept of random variables to account for unobserved effects caused by the lack of prevailing traffic data (Milton et al., 2007). Others relied on the simpler multinomial logit model (MNL) (Geedipally et al., 2013; Penmettsa and Pulugurtha, 2018). Statistically relevant factors from these studies include the number and radius of curves, pavement friction, interchanges/intersection density, rumble strips, barriers, underpass, on-ramps, area type (rural/urban), and weather. Traffic data used include daily traffic, average daily truck traffic, and truck percentages, however, prevailing traffic conditions were unavailable in these studies.

The hypothesis that high speeds and congestion might affect crash severity has led to the inclusion of prevailing traffic conditions in the set of explanatory variables used in discrete choice modeling

(Quddus et al., 2010). Crash and traffic data from London ring road freeway were used in an ordered response model to investigate the effect of congestion on road safety. Higher volumes were associated with less severe crashes, probably due to the lower speeds usually observed as demands increase. Other variables found to be statistically significant include curve radius, number of lanes, day of the week, lightning, road surface condition, number of vehicles involved, and year.

Many authors adjusted independent models for different classes of crashes. While Quddus et al. (2010) proposed a model for single-vehicle (SV) crashes only, Wu et al. (2014) conducted separate analyses for SV and multiple-vehicle crashes (MV) on rural two-lane highways in New Mexico. The authors concluded that the way the explanatory variables affect each model varies significantly for SV or MV crashes. Eleven variables were found to significantly affect driver injury severities for the MV model, albeit only seven of them could be related to the SV crash model.

Some studies focused on crash severity for one vehicle type. For motorcycles, Coutinho et al. (2015) used ordered response models to identify crash severity related factors for an urban environment, using data from the city of Fortaleza, Brazil. Chung et al. (2014) focused on Korean delivery-motorcycles. The results highlighted that, for motorcycle crashes, severity is more strongly related to human factors and nighttime conditions.

The question of what model type would best predict crash severity as a function of different sets of data prompted several researchers to focus on model comparison studies. Xie et al. (2009) used Bayesian inference into an ordered probit model (BOP), concluding that this approach could produce better results for small samples.

For rural single-vehicle crashes in Florida, the latent class logit (LCL) was compared to a traditional multinomial logit as a benchmark (Xie et al., 2012). Both modeling alternatives were able to capture the impact of each significant explanatory variable. Simpler models such as the ordered response models are preferable to more complex ones. From the 53 variables tested, 31 showed to be statistically significant at 95% confidence.

Other studies (Yasmin and Eluru, 2013; Çelik and Oktay, 2017) tested several ordered response frameworks to the classical unordered response framework such as the multinomial logit for crash

severity. Comparison of results indicated that the ordered model outperforms the multinomial logit. Similarly, Ye and Lord (2014) concluded that the unordered logit requires larger sample sizes to produce statistically reliable results, while the ordered, the smaller sample sizes.

2.3.2 SUMMARY

There is no consensus among the authors of previous studies regarding the superiority of any specific discrete choice method over the others, as the results of each study are dependent on the respective data used (Xie et al., 2012; Yasmin and Eluru, 2013; Çelik and Oktay, 2017; Ye and Lord, 2014). The ordered response model was therefore chosen for this work for it proved in previous work to provide reliable responses for various types of datasets, while keeping a simple structure. Also, ordered response models have the advantage of naturally following the logic of crash severity, in which the outcome is discrete and ranked in an order, in the case, fatalities, injuries and property-damage only crashes.

Differences between the severities of single-vehicle versus multiple-vehicle crashes were identified in many studies (Quddus et al., 2010; Wu et al., 2014). While some treated this characteristic as an additional variable, others advocate that separate models should be developed for each case, given the distinct nature and possible causes for each of these types of crashes. Some studies focused on investigating crash severity for specific types of vehicles, such as motorcycles (Chung et al., 2014; Coutinho et al., 2015).

Table 2.3 lists the most relevant studies from the literature, the size of the crash database used and period of study. The types of facilities included in each study is also shown, as well as the methods employed. The variables were grouped as crash characteristics, traffic and vehicle data, human factors, environment and geometry.

Most studies focused on generic roadways, and a few included intersection and control information as explanatory variables or by creating models for intersections. Few studies tested for differences between freeways, ramps and weaving segments.

Many studies raised the question whether speed or congestion (traffic density) influence crash severity, by analyzing indirectly related variables, such as area type, environment and visibility variables, estimates of speed at the time of the crash and annual average daily traffic (AADT). However, only Quddus et al. (2010) relied on a database that could combine prevailing traffic conditions to other roadway and environmental data, for one facility in the UK.

This research intended to analyze crash severity on several types of expressway segments in the light of a large database, covering an eight year period and counting on detailed prevailing traffic information and a comprehensive set of environment and geometry variables. As a limitation of this study, human factors were not available in the database.

Table 2.3 - Database size and characteristics of previous crash severity work.

		O'Donnell and Connor (1996)	Abdel-Aty (2003)	Milton et al. (2007)	Koekelman and Kweon (2008)	Quddus, Wang and Ison (2010)	Gedipally et al. (2013)	Coutinho et al. (2015)	Çelik and Oktay (2017)	Pennetsa and Pulugurtha (2018)	This Study
	Location	Aus.	FL	WA State	US A	London, UK	CA, ME and WA	CE, Brazil	Turkey	NC state	SP, Brazil
Facility	Road Segments	X	X	X	X				X	X	X
	Local Streets							X	X		
	Freeway Segments					X	X				X
	Ramps									X	X
	Weaves										X
	Intersections		X	X					X	X	
Crash	Count	18069	7891	22,568	NA	3998	NA	3232	11771	728801	21778
	Period	91	96-97	90-94	98	2003-06	2004-07	2004-11	2008-13	2011-13	2005-13
	Type	X			X				X		X
Method		OPM, OLM	OPM	MLM	OPM	OLM	MNL, NMNL	OPM, OLM	OLM	MNL	OLM
Traffic and Vehicle	AADT			X	X		X				X
	Average Speed					X					X
	Truck Percent			X							X
	Num. of Vehicle					X			X		X
	Vehicle Type	X	Car		X			Moto.	Heavy		X
	Vehicle Age	X			X						
	Private Vehicle							X	X		
Human Factors	Driver Age	X	X		X			X	X		
	Gender	X	X		X			X			
	Belt / Helmet	X	X					X			
	Alcohol Use	X	X		X						
	Speeding				X				X		
Environment	Lighting		X		X	X	X	X	X		X
	Weekend				X	X		X			X
	Weather		X	X		X			X		X
	Work Zones										X
	Urban/Rural		X				X				X
Geometry	Num. of Lanes					X			X		X
	Lane/Shoulder						X				
	Grades			X		X					X
	Curve		X	X		X	X				X
	Speed Limit			X		X					X
	Bridge/Tunnel									X	X
	Rumble Strips						X				
	Guardrails						X				X
Note:		MNL	Multinomial Logit Model				OPM	Ordered Probit Model			
		NMNL	Nested Multinomial Logit				OLM	Ordered Logit Model			
		MLM	Mixed Logit Model								

3 EFFECTS OF MOBILITY, SAFETY AND EMISSIONS ON SIGNAL TIMING OPTIMIZATION

The optimization of signal timings has been traditionally based on mobility performance measures only. There has been an increasing interest in utilizing signal timing optimization methods that can consider mobility, safety, and emissions measures simultaneously. The goal of this research is to develop and implement in an existing computational engine, the Highway Capacity Software (HCS), a signal timing optimization method that can consider mobility, safety, and emissions measures simultaneously. The HCS module Streets employs a genetic algorithm (GA) to optimize signal timings, using a single objective function traditionally based on mobility measures. As part of that deployment, this research investigated forms of incorporating safety and emissions measures in the HCS GA optimization, while keeping the simplicity of its single objective function. This required the development and implementation of new models to estimate safety and emission measures to supplement the existing HCS procedures that estimate mobility measures.

The introduction of the new models increased the complexity of the required inputs and the relationships between the inputs and outputs. This resulted in a large increase in input parameters to include factors related to geometric design, demand, control, optimization, and other features that impact the signal optimization results. It is important to understand the relative impacts of these features on the results of the signal timing optimization and the trade-offs between the resulting mobility, safety, and emission measures. This understanding will allow the designer, signal control engineer, and traffic analysts to consider these impacts when designing intersection geometry and when optimizing signal control. This research presents a method to optimize mobility, safety, and emission measures simultaneously and the application of this method in the HCS. It also proposes sensitivity analysis method to provide insight on the effects of 20 key variables on mobility, safety, emissions and their trade-offs.

The next section summarizes the optimization function components, focusing on crash prediction models for intersections, the emissions model used, and the selected sensitivity analysis method selected for this work. Next, the test scenarios and the optimization procedure are described. The

following section discusses the sensitivity analysis results. The last section summarizes the conclusions and limitations of the research, providing recommendations for future work.

3.1 METHODOLOGY

3.1.1 SIGNAL TIMING OPTIMIZATION

The existing procedure in the Streets module of the HCS incorporates a GA optimization algorithm to determine the offset, optimal cycle length, phasing sequence and splits. GAs are adaptive methods widely used to solve optimization problems using a direct analogy of natural behavior, working with a population of “individuals”, each representing a possible solution to a given problem. Individuals are composed by genes, each representing an input variable of the model. Each individual is assigned a “fitness score” according to how good a solution it produces is to the problem. The most highly fitted individuals are given opportunities to “reproduce”, by cross breeding with other highly fitted individuals (“crossover”) in the population. This produces new individuals as “offspring”, which share some features of their “parents”, but have a percent chance of “mutation”, reflected by a random change in one gene (parameter). The least fitted members of the population are less likely to be selected for reproduction, thus are weeded out gradually across “generations” of runs (Beasley et al., 1993).

Stevanovic et al. (2015) utilized GA in a signal timing optimization that simultaneously considers mobility, safety and emissions measures into a 3-dimensional objective function (Pareto Front). The study concluded that changes in the weights assigned to these measures did not produce significant differences in the results.

This study extended the GA optimization procedure of the HCS to consider all signal parameters in the optimization of the offsets, phasing sequence, split-phasing use, Dallas phasing use and cycle length. Cycles between 60 and 160 sec were allowed in the search for the optimal solution. The maximum value of 160 is 33% higher than the default recommended by the Highway Capacity

Manual (TRB, 2016) for major arterials with left turn phases in one street, allowing for a range of longer cycle lengths considered in the search region of the GA.

For the GA, the utilized population size was 10, with 30% crossover probability, a minimum 4.0% mutation probability, and 1% convergence threshold, as suggested by the HCS user guide. A maximum of 2,000 generations was used, which led to the convergence to the global optimum solution in most cases.

While the existing GA procedure in the HCS was kept, its objective function was modified to add the safety and emissions components to the optimization engine. Previous studies utilized a single objective function that combines multiple objectives in the GA optimization (Azar et al., 1999; Yan et al., 2013; Stevanovic et al., 2015). The fitness function for this study, which was based on the method proposed by Azar et al. (1999), is:

$$\text{Min} \left(w_{\text{delay}} \frac{f_{\text{delay}}(x) - \varepsilon_{\text{delay}, \text{good}}}{\varepsilon_{\text{delay}, \text{bad}} - \varepsilon_{\text{delay}, \text{good}}} + w_{\text{safety}} \frac{f_{\text{safety}}(x) - \varepsilon_{\text{safety}, \text{good}}}{\varepsilon_{\text{safety}, \text{bad}} - \varepsilon_{\text{safety}, \text{good}}} + w_{\text{emission}} \frac{f_{\text{emission}}(x) - \varepsilon_{\text{emission}, \text{good}}}{\varepsilon_{\text{emission}, \text{bad}} - \varepsilon_{\text{emission}, \text{good}}} \right), \quad 3.1$$

where w_{delay} , w_{safety} and w_{emission} are the weights to be used for the three performance measures; $f_{\text{delay}}(x)$ is the estimated overall delay (sec/veh), calculated by the HCM procedure; $f_{\text{safety}}(x)$ is the estimated total number of crashes over a five-year period; $f_{\text{emission}}(x)$ is the estimated emissions (g of gases); $\varepsilon_{i, \text{good}}$ is the minimum value for the performance measure i ; and $\varepsilon_{i, \text{bad}}$ is the maximum value for the performance measure i .

The HCM methodology was used for predicting delays. The crash prediction model by Turner et al. (2012) was selected. The emissions equations are presented in the following subsection.

3.1.2 EMISSIONS MODEL

This work used the model developed to estimate emissions by Chen et al. (2016) and Hadi et al. (2017). The model uses regression analysis, based on the outputs of the Environmental Protection Agency (EPA) MOVES, using information on the trajectories collected by the Federal Highway

Administration (FHWA) as part of the Next Generation SIMulation (NGSIM) program as inputs. More details about the development of the utilized model can be found in Hadi et al. (2017).

The general function for emission estimation is:

$$Emission = B_1 \times VMT + \frac{B_2}{avgspeed} + \frac{B_3 \times NumStops}{avgspeed} \quad 3.2$$

where B_1 , B_2 , and B_3 are coefficients of the independent variables. In summary, the best prediction models for various environmental measures are presented in Equation 3.3 to Equation 3.6:

$$CO(g) = 10.875 \times VMT + 3078.297 \times \frac{1}{avgspeed} + 1.470 \times \frac{NumStops}{avgspeed} \quad 3.3$$

$$NO_x(g) = 1.241 \times VMT + 185.490 \times \frac{1}{avgspeed} + 0.206 \times \frac{NumStops}{avgspeed} \quad 3.4$$

$$EC(joule) = 7.223 \times 10^6 \times VMT + 1.084 \times 10^9 \times \frac{1}{avgspeed} + 2.388 \times 10^6 \times \frac{NumStops}{avgspeed} \quad 3.5$$

$$CO_2Equi(g) = 519.1 \times VMT + 77950.6 \times \frac{1}{avgspeed} + 171.6 \times \frac{NumStops}{avgspeed} \quad 3.6$$

Note that the VMT is given in miles and the speed given in mph.

3.2 TEST SCENARIOS

A series of scenarios that consist of a base condition and variations to each input that might affect the outcome of the optimization procedure were used to test the proposed approach. At the end of each simulation, the following performance measures were obtained: (a) delays, to reflect mobility performance; (b) the number of predicted crashes to reflect safety; and (c) grams of gases, to quantify emissions.

3.2.1 BASE SCENARIO

A hypothetical three-intersection arterial was used for testing. This layout was chosen to allow for the optimization of arterial coordination parameters while keeping an acceptable degree of simplicity. All three intersections have the same configuration (Figure 3.1). Each arterial approach has two lanes with a 200-ft left-turn pocket and a shared right and through lane. There are no merges on the intersection exit sides. The side streets have two shared lanes. The intersection spacing is 1,300 feet. The length of all other segments is 1,000 feet. The speed limit of the arterial is 45 mph and that of the side streets is 30 mph.

The base traffic demands are also shown in Figure 3.1. The demand was determined iteratively after a series of tests using the HCS GA optimization, so that the level of service along the arterial ranged from C to E in the optimized base scenario, when considering all three objectives (mobility, safety and emissions).

Figure 3.1 - Configuration, demand and initial signal timing on testing arterial.

The parameters of the control setup used for the base scenario are:

- Coordinated arterial, with semi-actuated control and detectors present at each side street approach and protected left turn movements;
- Yellow time of 3 sec and red clearance of 2 sec for all movements;
- No Right-Turn on Red (RTOR);
- Exclusive left-turn phasing used for the main street and permissive only left-turn phasing used for the side streets (no split phasing).

The initial signal timing was the one optimized for the base scenario (Figure 3.1).

The following characteristics have significant impacts on the safety model only (Turner et al., 2012) and were set as described below:

- Intersection width of 110 feet: This is the distance between the stop bar on the approach and the beginning of the downstream intersection leg. This variable defines intersection size, as: (a) Small: depth of 82 ft or less and 1 or 2 lanes per approach; (b) Medium: depth between 82 ft and 131 ft and 2 lanes per approach; and (c) Large: depth larger than 131 ft and 3 or more lanes for at least two approaches.
- Median islands along the arterial;
- Overhead mast arm signal displays at each intersection;
- No parking along any of the links;
- No bus bays or bicycle facilities.

Other variables that affect safety, including degrees of saturation X and lost time per cycle L_C are calculated iteratively as intermediate results and are accounted for in the final iteration.

3.2.2 ALTERNATIVE SCENARIOS

For the creation of testing scenarios, three levels were considered per variable. Table 3.1 lists all variables included in the testing, along with their range and variation levels. The highlighted values are the inputs used in the base scenario. The following can be observed regarding the variations of these variables, as listed in Table 3.1:

- Volume Increments: The base volumes were decreased by applying multipliers of 0.9 and 0.8 to the base traffic flows entering from both the arterial and minor streets. As indicated earlier, oversaturated conditions were not considered in this study;
- Protected Left: When this variable is set as “Yes”, a permitted phase may also be suggested by the optimization engine. When it is set as “No”, no protected phases are allowed;
- RTOR: Two RTOR scenarios were considered. In one of them, the traffic volumes that could turn on red correspond to 25% of total right turning movements, while in the other, 50% could turn on red;
- Bicycle Facilities, Parkings and Bus Bays: Indicates the existence of one of those elements within 330 ft from the stop bar; and
- Weights: Three sets of scenarios were established for testing the weights in the optimization objective function. For each performance measure, these alternative schemes were tested: one-variable dominance (50/25/25, 25/50/25 or 25/25/50), or two-variable dominance (40/40/20, 40/20/40 or 20/40/40).

Table 3.1 - Variables included in the sensitivity analysis and variation range.

Domain	N°	Variable	Unit	Levels			Range
				1	2	3	
Design	1	Intersection spacing	feet	700	1300	1900	1200
	2	Intersection size		2	3	4	2
	3	Left turn pockets	feet	0	200	400	400
	4	Shared right-through lane		No	Yes		
	5	Shared left-through lane		No	Yes		
	6	Arterial posted speed	mph	35	45	55	10
Demand	7	Volume increments (multipliers)	veh/h	0.8	0.9	1	0.2
	8	Right turn percentages	%	5%	10%	15%	0.1
	9	Left turn percentages	%	5%	10%	15%	0.1
Control	10	Protected left		No	Yes		
	11	Right-turn on red (% right-turn flow)	%	0	25%	50%	0.5
Safety factors	12	Mast-arm signal display		No	Yes		
	13	Median island		No	Yes		
	14	Bicycle facilities		No	Yes		
	15	Bus bays		No	Yes		
	16	Parkings		No	Yes		
	17	Area type is residential		No	Yes		
Weight	18	Mobility		33/33/33	50/25/25	40/40/20	25
	19	Safety	%	33/33/33	25/50/25	20/40/40	25
	20	Emissions		33/33/33	25/25/50	40/20/40	25

3.3 SENSITIVITY ANALYSIS

A sensitivity analysis was undertaken to gain a better understanding of the interactions among the independent variables on safety, mobility, and emissions and the trade-offs between the three performance measures. Although the inputs are known, the magnitude of the effect of each input for different scenarios is unknown. To categorize the degree of importance of each input of the model, sensitivity analysis is recommended (Kuehl, 1999). Sensitivity analysis examines to what extent the outputs of a model depend on its inputs, by measuring the variation of the results as a function of different parameters (Trucano et al., 2006). This type of analysis also helps identify interactions between variables of the model (Saltelli et al., 2004).

Different classes of sensitivity analysis methods exist for different purposes (Santner et al., 2003). Local sensitivity analysis methods are used to identify trends for a specific variable within a given sample space, while global sensitivity analysis considers the entire population.

In this work, 20 variables that could affect the model outcome were identified, eleven to be tested at two variation levels and nine to be tested at three variation levels. In order to cover all combinations of values for the 20 inputs, $(39) \times (211) = 4 \times 10^7$ scenarios would have to be generated, which was unfeasible for the available computational resources. In this case, when the search area is too large to be tested in its entirety, screening methods are recommended (Saltelli et al., 2000). The best-known screening method is the classical *Ceteris Paribus*, in which one factor is varied within a range at a time, with all other variables held constant. A statistically significant sample is calculated for the number of scenarios to be constructed. Since only one variable is changed at a time, this method has limitations regarding the identification of interactions between variables, and restrictions in terms of possible scenarios.

To remedy this situation, the Factorial Effect Method (FEM) was proposed by Morris (1991), further developed by Campolongo et al. (2007) and successfully applied in another transportation engineering study (Nunes, 2012). It is based on the concept of factorial effects – $d_i(x)$, defined for each input and measured analogously to price elasticities:

$$d_i(x) = \frac{MOE(x_k) - MOE(x_{k-1})}{\Delta_i} \quad 3.7$$

where $d_i(x)$ is the factorial effect of the variable i ; $MOE(x_k)$ are the performance measures for a vector on input data k ; and Δ_i is the variation of the factor i between vector x_k and x_{k-1} .

Sampling consists of assembling an input matrix in a procedure called “trajectory” generation. The first row of the matrix reflects an initial vector of inputs, generally corresponding to a base scenario. The second row is defined by changing one variable by Δ_i . Likewise, the third row of the matrix is constructed by varying another variable that hasn’t been changed in any previous row. This is done successively, never changing the same variable twice, until all inputs have been shifted. The order

in which each variable is changed is random for each trajectory, so that no two matrices are the same. A trajectory j of three variables model would be represented as:

$$j = \begin{vmatrix} x_1 & x_2 & x_3 \\ x_1 & x_2 + \Delta_2 & x_3 \\ x_1 + \Delta_1 & x_2 + \Delta_2 & x_3 \\ x_1 + \Delta_1 & x_2 + \Delta_2 & x_3 + \Delta_3 \end{vmatrix} \quad 3.8$$

For each trajectory, each factor is tested only once. To have statistically significant results, multiple trajectories must be constructed. From the previous experience by Morris (1991), a minimum of 10 trajectories is sufficient for a reliable analysis, while more than 30 trajectories didn't add more confidence to the results by Nunes (2012).

When all vectors from all trajectories j are simulated, the factorial effects $d_{i,j}(x)$ of all variables are compiled, forming the distribution F_i . The absolute values of $d_{i,j}(x)$ form the G_i distribution. Based on those distributions, sensitivity measures are calculated for each variable:

$$\mu_i = \sum_{j=1}^r \frac{d_{i,j}}{r} \quad 3.9$$

$$\sigma_i = \sqrt{\sum_{j=1}^r \frac{(d_{i,j} - \mu_i)^2}{r}} \quad 3.10$$

$$\mu_i^* = \sum_{j=1}^r \frac{|d_{i,j}|}{r} \quad 3.11$$

where r is the number of trajectories j .

The main advantages of FEM are:

- Sampling: it produces randomized samples of variables so that, differently from the pure Ceteris Paribus method, different combinations of variables are tested. This results in distribution of effects that allow the researcher to verify not only the mean effects of each variable, but also variance, biases and normality assumptions;

- By using a measure analogous to price elasticities, different variables are brought to a similar unit and can be compared and ranked in order of importance to the model.

3.3.1 APPLICATION

The application of the FEM is based on the construction of trajectories, composed by vectors of input data. Each trajectory is represented by a matrix with 20 columns (number of variables) and 21 rows, in which the first row reflects the base scenario and all other rows represent variations of each variable within one of the levels (Table 3.1). Table 3.2 was developed to generate random trajectories. From the base scenario, each consecutive vector is composed by varying one single input (highlighted), to one of the possible levels. Selected levels and the order in which the variables are assorted are randomly defined.

A second matrix was also constructed, containing unfeasible or unreasonable combinations of inputs. Examples of these are increasing left-turn pockets while shared left through lanes are in use, or the use of exclusive right and left turns when the arterial has two lanes, which would imply there are no through lanes. The trajectory generator (Table 3.2) then constructs random scenarios until all vectors are valid.

Table 3.2 – Random trajectory generator.

	Intersection spacing	Intersection Size	Left turn pockets	Shared right turn	Shared left turn	Posted Speed	Volume increments	Right turn percent	Left turn percent	Protected Left	Right-turn on red	Mast-arm display	Median island	Bicycle facilities	Bus bays	Parkings	Area is residential	Mobility	Safety	Emissions
Base Vector	2	2	2	2	1	2	3	2	2	2	1	2	2	1	1	1	1	1	1	1
Median island	2	2	2	2	1	2	3	2	2	2	1	2	1	1	1	1	1	1	1	1
Left turn percent	2	2	2	2	1	2	3	2	3	2	1	2	1	1	1	1	1	1	1	1
Right turn percent	2	2	2	2	1	2	3	1	3	2	1	2	1	1	1	1	1	1	1	1
Shared right turn	2	2	2	1	1	2	3	1	3	2	1	2	1	1	1	1	1	1	1	1
Mast-arm display	2	2	2	1	1	2	3	1	3	2	1	1	1	1	1	1	1	1	1	1
Right-turn on red	2	2	2	1	1	2	3	1	3	2	3	1	1	1	1	1	1	1	1	1
Bicycle facilities	2	2	2	1	1	2	3	1	3	2	3	1	1	2	1	1	1	1	1	1
Posted Speed	2	2	2	1	1	3	3	1	3	2	3	1	1	2	1	1	1	1	1	1
Intersection spacing	3	2	2	1	1	3	3	1	3	2	3	1	1	2	1	1	1	1	1	1
Volume increments	3	2	2	1	1	3	2	1	3	2	3	1	1	2	1	1	1	1	1	1
Bus bays	3	2	2	1	1	3	2	1	3	2	3	1	1	2	2	1	1	1	1	1
Safety	3	2	2	1	1	3	2	1	3	2	3	1	1	2	2	1	1	1	2	1
Protected Left	3	2	2	1	1	3	2	1	3	1	3	1	1	2	2	1	1	1	2	1
Shared left turn	3	2	2	1	2	3	2	1	3	1	3	1	1	2	2	1	1	1	2	1
Left turn pockets	3	2	3	1	2	3	2	1	3	1	3	1	1	2	2	1	1	1	2	1
Area is residential	3	2	3	1	2	3	2	1	3	1	3	1	1	2	2	1	2	1	2	1
Parkings	3	2	3	1	2	3	2	1	3	1	3	1	1	2	2	2	2	1	2	1
Emissions	3	2	3	1	2	3	2	1	3	1	3	1	1	2	2	2	2	1	2	2
Intersection Size	3	1	3	1	2	3	2	1	3	1	3	1	1	2	2	2	2	1	2	2
Mobility	3	1	3	1	2	3	2	1	3	1	3	1	1	2	2	2	2	2	2	2

Each scenario (vector) was simulated, using the levels shown in Table 3.2 and their correspondent values shown in Table 3.1. For each run, the resulting performance measures for mobility, safety and emissions were recorded. The factorial effect of each variable is then calculated as per Equation 3.7, and the three statistics given in Equations 3.9-3.11 were calculated. The higher the μ^* , the higher the indication that a variable has a significant impact on a performance measure.

3.3.2 RESULTS

Figure 3.2 graphically shows the results obtained from a total of 30 trajectories (30 trajectories $\times$ 20 variables = 600 individual scenarios), in terms of the distribution of the factorial effects across

all trajectories. A positive sign indicates a positive relationship between the variable and the performance measure (dimensionless, analogously to price elasticities). For instance, an increase in traffic volume is associated with increases in delay, crashes and emissions.

For practical purposes, variables with a negative sign for the mobility and emissions indicate improvements in all performance measures (less overall delay, emissions and fewer crashes). For example, for the intersection size, the higher the number of lanes, the lower the delay and emissions. The degree of improvement for these variables oscillated between scenarios, as reflected by the dispersion of the results (higher variability).

For the binary variables, the “yes” value is associated with 1. As an example, the use of a protected left phase increased the delay at a rate ranging from 2.5 to 7 sec/veh (median of 5). Demand level variables, such as the volume multipliers, are directly related to all performance measures. For most scenarios, through movements were the critical ones, and the use of a lane for shared turns and through movements enhanced mobility and emissions, albeit often at the expense of safety.

The emission results followed a similar trend to that observed for the mobility domain. For instance, RTOR leads to a slight reduction of emissions that could be considered statistically more significant than at least ten other variables, at 95% confidence.

Intersection spacing was found to increase emissions more than other variables. This is expected, since longer segment lengths result in more driving and higher emissions.

Giving more weights to specific performance measures (variables Mobility, Safety and Emissions, respectively) had a limited effect on the results. This suggests that the optimum solutions attained by the proposed algorithm are only marginally affected by the user-defined weights w_{delay} , w_{safety} and $w_{emission}$.

Figure 3.2 - Factorial effects for each input parameter and performance measure.

3.3.3 DISCUSSION ON THE TRADE-OFF BETWEEN MEASURES

While the impact of some variables, such as demand levels, is straightforward for all performance measures, for many others the factorial effects deviate around 0. This suggests that the influence of that particular variable on a specific performance measure depends on the combination of other variables (i.e., the particular scenario analyzed), and a trade-off among different measures may exist.

For example, for the case of permissive versus exclusive left turns, two different effects were identified: the use of permissive left turns creates an additional conflict that increases the risk of left-turn-against crashes. However, depending on the traffic flows per approach/movement, this scheme has the potential to reduce the degrees of saturation, phase complexity and cycle time, improving the intersection safety. Due to this trade-off, depending on the scenario, there were overall increases/decreases in the number of crashes. Larger intersections were associated with more angle and rear-end crashes when no left-turn pockets were present.

Shared Left Turns increase the number of predicted crashes when used to replace an exclusive left turn lane with a pocket. On the other hand, for the case of large intersections (4+ lanes on the arterial) with short/no pockets provided for the left turn, the use of shared left turns was found to be a better design. The effect of shared right turns versus exclusive right turns is simpler. In this case, the use of shared right turns decreases delay at the cost of safety for all simulated scenarios.

Emissions are a function of three main factors: distance-vehicles travelled (VMT), number of stops, and average speed. The intersection size was defined by the number of lanes, which in turn affects the road width. Depending on the scenario (combination of variables), better intersection performance is obtained with the addition of extra lanes and that causes a decrease in the number of stops and an increase in the average speed, reducing emissions. On the other hand, when the extra capacity is unnecessary because of lower degrees of saturation, the additional lanes result in extra distance to be covered between the approaches (road width), increasing VMT and, thus, emissions.

3.3.4 RANKING OF INFLUENCING VARIABLES BY MEASURE

Although the central and dispersion measures shown in Figure 3.2 provides some insight on the magnitude of the impact of each variable, the dispersion of some results suggests that one-to-one comparisons of variables with similar effects is not straightforward. For that reason, a multiple pairwise comparison statistical test was performed. The analysis used the absolute factorial effects (μ^*) of each variable, which is considered to be the most statistically significant measure in the method (Campolongo et al., 2007).

In a preliminary analysis, it was noted that the distribution of μ^* did not follow a normal (Gaussian) pattern. This happens because the interaction between variables generates scenarios for which the results diverge from the mean unexpectedly, as discussed in the previous subsection. For that reason, the non-parametric Kruskal-Wallis test was chosen, which accounts for the frequency that each variable assumes a specific position in the overall ranking. The results for the test, at the 95% confidence level, are shown in Table 3.3. The variables are ranked in order of importance, per the Kruskal Wallis test criteria, according to the number of times each one of them figured among the most relevant to the model outcomes.

The higher value of the test statistic H as compared to the critical χ^2 indicates that H_0 can be rejected for all performance measures, and at least two variables would differ from each other. To specify which ones, pairwise comparisons were made. The table shows which groups of variables could be ranked higher relatively to the others.

Table 3.3 – Kruskal-Wallis test for all variables and performance measures.

Delay		Number of Crashes		Emissions	
Critical χ^2	H	Critical χ^2	H	Critical χ^2	H
42.6	271	42.6	157	42.6	421
Variable	Rank	Variable	Rank	Variable	Rank
Protected Left	1	Volume increments	1	Intersection spacing	1
Intersection Size	1	Left turn pockets	1	Intersection Size	1
Volume increments	1	Intersection Size	1	Shared right-through	1
Shared right-through	1	Mast-arm signal display	1	Shared left-through	1
Intersection spacing	2	Median island	1	Volume increments	1
Shared left-through	2	Shared right-through	1	Left turn percentages	2
Area type is residential	2	Bus bays	1	Right turn percentages	2
Left turn percentages	2	Parkings	1	Arterial Posted Speed	3
Mobility	2	Right-turn on Red	1	Protected Left	4
Right turn percentages	2	Shared left-through	1	Right-turn on Red	4
Emissions	2	Bicycle facilities	2	Area type is residential	4
Safety	3	Area type is residential	3	Mobility	5
Arterial Posted Speed	3	Protected Left	4	Emissions	6
Right-turn on Red	3	Arterial Posted Speed	4	Safety	6
Left turn pockets	4	Left turn percentages	4	Bicycle facilities	6
Median island	4	Mobility	4	Median island	6
Mast-arm signal display	4	Emissions	5	Mast-arm signal display	6
Bicycle facilities	4	Right turn percentages	5	Left turn pockets	6
Bus bays	4	Intersection spacing	5	Bus bays	6
Parkings	4	Safety	5	Parkings	6

3.3.5 IMPACT OF WEIGHT SCHEMES ON SIGNAL TIMING

The final stage of this research assessed the impacts of weight schemes on the signal timings. Using the base scenario (Figure 3.1), three additional scenarios were generated, giving all the weight to one performance measure at a time.

The resulting performance per intersection and weight scheme are shown in Table 3.4. As expected, providing all weight to the mobility measure resulted in the best overall operational performance. When the weight is assigned to either safety or mobility, the movements with the highest traffic

demands (usually the main street) have longer green times and the minor streets and left turn movements have the shortest green times.

The phasing sequence was kept unaltered by the algorithm in all cases, but the cycle length varied significantly with the changes in the weights on the three measures: 65 sec for the 100% mobility weight, 160 sec for the safety scheme and 150 sec for the emissions scheme.

The selection of shorter cycles for optimal mobility performance can be explained by the lower degree of saturation of the intersection movements. Congested conditions were not examined in this work and thus longer cycles were not selected to provide more capacity to the movements. Regarding safety, longer cycle lengths were selected since longer green times allow more gaps in the opposing through traffic for permitted left turn movements and thus reduce the potential for accepting shorter gaps that potentially result in crashes. Also, by reducing the number of times of phase changes, longer cycles further reduce rear-end, left-turn and loss-of-control crashes that mainly occur during the phase change periods. Regarding emissions, longer cycles tended to reduce the number of stops, which is related to the increase in all types of emissions.

Table 3.4 – Results for different optimization schemes, per intersection.

Approach Movement	EBL	EBT	EB R	WB L	WB T	WB R	NB L	NB R	SBL	SBR
Demand (veh/h)	108	860	108	76	608	76	20	20	20	20
Intersection 1	Mobility Optimization									
	Level of Service (LOS)	C	C	C	C	B	B	C	C	C
	Intersection Delay / LOS		21.3					C		
	Num. of Crashes (5 years)		6.25					480.6		
	Safety Optimization									
	Level of Service (LOS)	E	B	B	E	A	A	E	E	E
	Intersection Delay / LOS		30					C		
	Num. of Crashes (5 years)		4.55					470.2		
	Emissions Optimization									
	Level of Service (LOS)	F	B	B	E	A	A	E	E	E
	Intersection Delay / LOS		28.6					C		
	Num. of Crashes (5 years)		4.65					461.0		
Intersection 2	Demand (veh/h)	90	720	90	90	720	90	20	20	20
	Mobility Optimization									
	Level of Service (LOS)	C	A	A	C	A	A	C	C	C
	Intersection Delay / LOS		9.9					A		
	Num. of Crashes (5 years)		5.2					465.8		
	Safety Optimization									
	Level of Service (LOS)	F	A	A	E	A	A	E	E	E
	Intersection Delay / LOS		21.8					C		
	Num. of Crashes (5 years)		4.2					470.2		
	Emissions Optimization									
	Level of Service (LOS)	E	A	A	E	A	A	E	E	E
	Intersection Delay / LOS		21.2					C		
	Num. of Crashes (5 years)		4.3					458.3		
Intersection 3	Demand (veh/h)	76	608	76	108	860	108	20	20	20
	Mobility Optimization									
	Level of Service (LOS)	C	A	A	C	A	A	C	C	C
	Intersection Delay / LOS		13.9					B		
	Num. of Crashes (5 years)		5.15					465.8		
	Safety Optimization									
	Level of Service (LOS)	F	B	B	E	B	B	E	E	E
	Intersection Delay / LOS		26.9					C		
	Num. of Crashes (5 years)		4.25					474.3		
	Emissions Optimization									
	Level of Service (LOS)	F	A	A	E	B	B	E	E	E
	Intersection Delay / LOS		23.4					C		
	Num. of Crashes (5 years)		4.35					458.3		

3.4 FINAL REMARKS

A methodology was developed and implemented in a computational engine (the Highway Capacity Software – HCS), employing a genetic algorithm now capable of optimizing signal timing and coordination on arterials, allowing the optimization to be based not only on mobility but also safety and emissions. The resulting methodology is available to the users as part of the HCS – Streets module.

To attain this goal, an emissions model and a set of crash prediction equations were adapted and incorporated to the HCS existing single optimization function. The selected safety model has the advantage of considering geometry and control aspects, while its multiplicative structure allows for local calibration of existing parameters and future inclusion of additional factors, such as HSM Crash Modification Factors, the number of phases, maximum greens or the presence of countdown timers, for example.

This research presents a new method, the results from its implementation and sensitivity analysis conducted to provide insight on the effects and order of relevance of 20 key variables on the model's outcomes and the associated trade-offs between mobility, safety, and emission. This insight will help the designer, signal control engineer, and traffic analyst when designing intersection geometry and signal control.

By using the Factorial Effects Method (FEM), it was concluded that, while the impacts of groups of variables on the mobility, safety, and emission measures are noticeably more significant than others, the effect of several of them is highly dependent on the combination with other variables, so that the individual degree of importance of each variable to the model cannot be fully ranked, at the 95% confidence.

Demand level variables and the size of the intersection, defined by the number of lanes on the arterial, were found to largely affect all three performance measures. Large intersections with no pockets were associated with a much higher number of angle and rear-end crashes. This effect can be mitigated with either the use of exclusive left turn pockets or by allowing the left-turn movement on this lane to be shared with the through movement.

Protected left turns increase delay, but only improve safety when left turn volumes are significant, imposing extra delay and crash risk. Right-turn on red were associated with a few more crashes, while slightly improving the overall mobility and reducing emissions.

Weight schemes for the performance measures had a limited impact on the optimal results, suggesting that the single objective optimization algorithm converges for a similar optimum solution regardless of the user-defined weights.

A limitation of this study is that some of the safety-related variables could not be fully tested for their effect on other performance measures. The presence of bus bays, parking or bicycles facilities were analyzed in terms of safety impacts only, due to the large variability and uncertainties of operation effects of these variables. Finally, future research should incorporate pedestrian mobility and safety to the optimization method.

4 CRASH AND TRAFFIC DATABASE FOR EXPRESSWAYS

This study used a database built from different sources, covering a period from 2005 to 2013, comprising 81,484 crashes and observations of traffic volumes and speeds by vehicle class collected at 5 to 15-minute intervals by 187 traffic monitoring stations on 13 toll expressways in São Paulo state, Brazil (Figure 4.1). These expressways consist of 634 km and 818 km of directional freeways and divided multilane highways respectively. They serve over 22 million people in their vicinities.

Figure 4.1 – Studied freeways, multilane highways and main urban areas.

All traffic monitoring stations are located on segments with 2 to 5 lanes with a 3.5 m (11.5 ft) width, and 2.5 m (8.2 ft) wide shoulders. All multilane highway segments are divided and have no at grade crossings or traffic signals, but can be accessed by regular driveways (not necessarily ramps). Shoulders can be used by cyclists and pedestrians in paved roads in the absence of bike facilities, particularly in urbanized areas. Automobile posted speeds range from

80 to 120 km/h (50 – 75 mph); the limits for trucks are lower on most highways, similarly to the policies in force in the states of California, Indiana, Montana, Oregon and Washington (NMA, 2018).

The crash database includes the type, severity, number, and types of vehicles involved. The date, time and location of each crash are also specified. This information includes not only the milepost, but also whether the crash occurred on the main lanes, shoulders, service roads, intersections, tunnels, or toll plaza booths.

The data were supplemented by information regarding work zones, periods with special operation schemes and enforcement. This made it possible to: a) exclude periods when road work zones were active, or when police or automated enforcement were in operation; and b) with the support of historical Google Earth images, identify major changes on the infrastructure (opening or closure of ramps, number of lanes etc.), helping the classification and characterization of each segment over time.

Each crash was assigned to a segment monitored by a corresponding traffic station for the same period. Those segments were classified as basic segments, weaves and segments in the influence area of on-ramps or off-ramps, as defined in the Highway Capacity Manual (TRB, 2016). Segments of the studied highways that could not be classified in one of those types were not included in the study. Likewise, crashes outside the boundaries of monitored segments, within the influence of toll plazas or during excluded periods were excluded from the data. As a result, 1039 km of the original 1452 km of highways and 21,969 of 81,484 crashes were included in this study.

Table 4.1 shows the frequency of roadway characteristics that might affect operations and safety, the total exposure (mi of veh-km) and the number of crashes for the study period. While the relative length of urban segments is lower compared to rural segments, the number of stations for both categories is similar. In some cases, crash occurrence may be associated to higher exposure. In others, crash frequency is high in contrast to the exposure. In example, five control stations on weaves cover 0.5% of the total length, corresponding to 1% of the exposure and 2% of the number of crashes.

Table 4.1 – Frequency of main road characteristics, number of crashes and exposure

Variable	Categories	Traffic Monitoring Stations	Freq. (%)	Length (km)	Freq.	Exposure (veh-km/10 ⁶)	Freq. (%)	Crashes	Freq. (%)
Road Type	Freeway Segment	65	34.8	425.3	44.4	8917	43.8	9306	42.4
	Multilane Highway	87	46.5	512.9	53.5	10802	53.1	11621	52.9
	OnRamp	18	9.6	10.2	1.1	329	1.6	413	1.9
	OffRamp	12	6.4	5.4	0.6	102	0.5	186	0.8
	Weaving	5	2.7	4.4	0.5	206	1.0	443	2.0
Area	Rural	94	50.3	691.7	72.2	12627	62.0	11179	50.9
	Urban	93	49.7	266.5	27.8	7729	38.0	10790	49.1
Num. of Lanes	2	99	52.9	567.8	59.3	8722	42.8	10268	46.7
	3	59	31.6	302.4	31.6	6680	32.8	7663	34.9
	4	25	13.4	78.1	8.2	4429	21.8	3438	15.6
	5	4	2.1	9.8	1.0	525	2.6	600	2.7
Posted Speed	80	16	8.6	33.9	3.5	567	2.8	1245	5.7
	90	4	2.1	12.4	1.3	88	0.4	68	0.3
	100	80	42.8	212.3	22.2	8059	39.6	9528	43.4
	110	31	16.6	237.9	24.8	1688	8.3	1581	7.2
	120	56	29.9	461.8	48.2	9949	48.9	9526	43.4

Figure 4.2 shows the distribution of severity among the three levels considered in this study, for the whole database [81,484 crashes, Figure 4.2 (a)] and for the used dataset [21,969 crashes, Figure 4.2 (b)]. The crashes were classified as: Property-damage only (PDO), Injury (I), Fatal (F), for SV crashes, MV crashes MV and pedestrian and bicycle crashes (P&B). The percentages shown on the top reflect the severity distribution within each class, while the numbers in brackets represent the overall proportions of each crash class and severity relative to the total.

Fatal crashes correspond to 2.2% of the total. Note that there are no PDO crashes for the P&B class, and fatalities for these can be as high as 39.1% of the total number of fatal crashes.

Figure 4.2 – Distribution of crashes by class and severity.

(a) Multiple-vehicle crashes

(b) Single-vehicle crashes

All variables of the database are shown in Table 4.2. For variables with multiple categories, binary variables were created, while continuous variables were kept unchanged. In addition to these variables, the year of each crash occurrence was included, in order to avoid any bias from temporal trends in the data.

The proportion of each crash class per category is also shown, allowing for the assessment of each class can be more associated to a particular characteristic category. example, regarding crash type, runoffs and hit animal are typical single-vehicle crashes, while rear-end, angle, sideswipe and head-on must involve two or more vehicles.

The discrete variables can be described as:

- Area Type: urban areas include urban and suburban abutting land of cities with more than 5,000 inhabitants, as well as industrial areas.
- Crash Types: collision with objects, animals and run-offs are single-vehicle crashes, while head-on, sideswipe, angle and rear-end crashes may involve two or more vehicles.
- Weekend: weekend days comprises Saturday and Sunday, not including Friday nights. Other days are labeled as weekdays.
- Nighttime: the sunrise and sunset times were obtained for each day of the year from data by the U.S. Naval Observatory, available at IAG/USP (2018), so that each crash could be assigned to the daytime or nighttime periods accounting for the effect of the seasons.
- Visibility: perceived overall visibility by the driver and officer during crash reporting.
- Weather: drizzle was defined as rain that is sufficient to cause the pavement to become wet but does not impair visibility. The category “rain” affects both the pavement surface and visibility, while fog does not affect the surface and heavily impacts visibility.
- Vertical Alignment: all crashes happening in grades steeper than 3% were assigned the categories “upgrades” or “downgrades”, according to the direction of traffic.
- Horizontal Alignment: defined according to the segment where the crash happened, being curvature (deg/km) used as a measure. Values under 100 deg/km were associated to straight segments. From 100-500 deg/km, segments were labeled as smooth curved, while segments with a curvature value above 500 deg/km were considered sharp curves.
- Segment Type: the presence of ramps or weaves was considered as an explanatory variable and their influence areas were defined per the HCM criteria (TRB, 2016).
- Bridge: this variable refers to bridge segments themselves, not including underpasses.
- Guardrails: this variable was assigned to all crashes in which vehicles hit a median or roadside barrier or metallic guardrail.
- Road works: this included warning and speed reduction areas from work zones.

Table 4.2 – Discrete variables included in the database.

Variable	Categories	Frequency		Proportion by crash class		
		Total	Relative	MV	P&B	SV
Area Type	Rural	5722	44.8%	52.9%	2.3%	44.8%
	Urban	7049	55.2%	65.5%	3.0%	31.5%
Crash Type	Other	1097	8.6%	14.6%	0.0%	85.4%
	Collision w/ object	779	4.3%	61.5%	0.0%	38.5%
	Hit Animal	229	1.0%	3.5%	0.0%	96.5%
	Run-off	4340	23.6%	8.2%	0.0%	91.8%
	Head-on	131	1.3%	100.0%	0.0%	0.0%
	Sideswipe	1978	16.4%	100.0%	0.0%	0.0%
	Angle	257	2.6%	100.0%	0.0%	0.0%
	Rear-end	4340	32.4%	100.0%	0.0%	0.0%
Weekend	No	9101	71.3%	49.5%	3.1%	47.3%
	Yes	3670	28.7%	63.2%	1.5%	35.3%
Nighttime	No	7276	57.0%	55.4%	4.2%	40.3%
	Yes	5495	43.0%	61.9%	2.5%	35.6%
Visibility	Good	10912	85.4%	48.6%	3.9%	47.5%
	Average	1762	13.8%	56.0%	0.0%	44.0%
	Poor	25	0.2%	37.5%	6.3%	56.3%
Weather Condition	Good	10973	85.9%	58.2%	0.3%	41.5%
	Drizzle	294	2.3%	37.0%	7.4%	55.6%
	Rain	1396	10.9%	61.9%	2.9%	35.2%
	Fog	27	0.2%	55.1%	2.7%	42.2%
Vertical Alignment	Level	2992	23.4%	59.6%	3.1%	37.3%
	Upgrade	2156	16.9%	65.7%	2.7%	31.6%
	Downgrade	2512	19.7%	43.8%	1.3%	54.9%
Horizontal Alignment	Tangent	8486	66.4%	54.7%	3.7%	41.7%
	Smooth Curve	1390	10.9%	62.3%	2.7%	35.0%
	Sharp Curve	625	4.9%	64.0%	2.2%	33.8%
Segment Type	Segment	12209	95.6%	56.9%	3.1%	40.0%
	Ramp	305	2.4%	61.9%	1.2%	36.9%
	Weave	257	2.0%	59.9%	2.7%	37.4%
Bridge	No	12749	99.8%	63.6%	0.0%	36.4%
	Yes	22	0.2%	65.7%	3.0%	31.3%
Guardrail	No	10882	85.2%	9.1%	0.0%	90.9%
	Yes	1402	11.0%	59.7%	3.0%	37.3%
Road works	No	6451	50.5%	57.6%	2.3%	40.1%
	Yes	2494	19.5%	14.6%	0.0%	85.4%

Other variables were treated as continuous variables, as shown in Table 4.3. The number of vehicles involved in a MV crash, the posted speed and number of lanes used integer values,

while other traffic variables were defined by real numbers. The density variation variable was calculated as the difference between the density right prior to the crash and at the time interval immediately before.

Table 4.3 – Continuous variables included in the database.

Variable	Mean	Std. Deviation	Min	Max
Number of Vehicles	1.75	0.83	1	16
Traffic Density (pc/km/lane)	9.59	6.98	0	32
Average Speed (km/h)	89.18	16.02	0	174
Percent of Trucks	22.87	15.20	0	100
Density Variation (pc/km/lane)	-1.11	4.98	-32	66
Posted Speed (km/h)	107.65	11.46	80	120
Number of Lanes	2.76	0.82	2	5

Note that, information such as substance abuse, age or gender of the drivers involved in the crash were not available.

5 RELATIONSHIP BETWEEN CRASH RATE AND OPERATIONS ON EXPRESSWAYS

Studies that investigated crashes as a function of hourly flow rates rather than as a function of AADT focused on urban freeway segments, and generally concluded that crash rates are higher for both low and high traffic flow conditions, reaching a minimum when traffic density is moderate (Frantzeskakis and Iordanis, 1981; Ceder, 1982; Chang et al., 2000; Martin, 2002; Zhou and Sisiopiku, 2014; Harwood et al., 2013; Potts et al., 2014). Those studies often counted on a limited amount of data and could account for the influence of a restricted number of road and environmental characteristics. Some rely on data from a single facility, that provides valuable insight but is insufficient to establish broader generalizations. Most studies couldn't distinguish crashes that happened on different road types, such as basic segments, weavings or ramps. Finally, few data from rural highways were available.

This work was motivated by the availability of a large and diverse database containing about 35 million traffic observations and over 20,000 crashes from urban and rural multilane highways, freeway segments, merges, diverges and weaves with different geometry and environmental conditions, that allows for in-depth investigation of the relationship between traffic density and crash frequency for different situations.

The objectives of this stage of the research were to:

1. Model the relationship between traffic density and crash rate in expressways and compare the results to previous findings in the literature;
2. Develop a novel model that incorporates significant road and traffic characteristics, capable of predicting hourly crash rates based on several roadway and traffic variables; and
3. Discuss the implications of the research findings for highway design and operations.

The next section of this chapter describes the metrics and the relationship between density and crash rates derived for the whole dataset, which confirmed trends found in previous studies. The fourth section explains the methodology used to build the model proposed in this paper, which was comprised of selecting statistically significant variables and fitting equations by using regression modeling. The chapter concludes with an analysis on the consistency of the

proposed method to traditional crash prediction methods, along with a discussion on the implications of the findings and final remarks.

5.1 RELATIONSHIP BETWEEN DENSITY AND CRASH RATE

This research uses traffic density (pc/km/lane) to model the relationship between crash rate and traffic operational quality. Density (D) was calculated for each 5 to 15 minute interval as:

$$D = \frac{v}{S} \quad 5.1$$

where S is the average speed (km/h) measured in the field at the stations indicated, and v is the equivalent flow in passenger cars per lane per hour. Truck volumes were converted to passenger car flows by using equivalents calculated by a previous study (Piva, 2015) from a sample of the same database.

The database was then divided in density bins of 3 pc/km/lane. Each bin was assigned the total number of crashes, considering the density value of the time interval immediately prior to the crash, as well as the corresponding exposure, which is the traffic volume during the same interval multiplied by the segment length (veh-km). The crash rate (crashes per million veh-km) for each density bin b is then calculated as:

$$Crash\ Rate_b = \left[\frac{\sum_b \sum_i^n Crashes_i}{\sum_b \sum_i^n (L_i \times V_i)} \right] \times 10^6 \quad 5.2$$

where $Crashes_i$ and V_i are the total number of crashes and the total number of vehicles respectively, on segment i within the density b ; L_i is the length of each segment i ; n is the total number of segments.

The aggregate result for the whole dataset confirmed findings from previous studies (Frantzeskakis and Iordanis, 1981; Ceder, 1982; Chang et al., 2000; Zhou and Sisiopiku, 2014; Harwood et al., 2013; Potts et al., 2014), and is shown in Figure 5.1. The relationship can be roughly described by a cubic polynomial, with the minimum rate of crashes at LOS B and C. For the same LOS, the proportion of FI and PDO crashes are similar. For more congested conditions, however, PDO crashes become more prevalent.

Figure 5.1 – Relationship between traffic density and crash rate, by severity.

Figure 5.2 shows the relationships by crash class. As in previous studies, the combination of the higher frequency for single vehicle crashes at low density and the rise of multiple-vehicle crashes at higher densities explains the U-shape form of the relationship. Crashes with pedestrians and bicycles follow a similar relationship, with noticeable higher frequencies at low density. A better understanding of this type of crash, which accounts for 39% of fatal crashes in the database, is recommended, and would require the incorporation of additional factors, such as pedestrian and bicycle volumes.

Figure 5.2 – Relationships between traffic density and crash rate, by class.

Based on this analysis, it can be concluded that for this database the overall relationship between density and crash rate is consistent with the relationship found in other similar databases in the US.

Differently from previous studies however, a slight increase of single vehicle-crashes beyond the boundary of the level of service E was observed. To understand this phenomenon, the graph shown in Figure 5.3 was constructed, comparing the proportion of single-vehicle crash types for undersaturated (LOS A-E) and oversaturated (LOS F) conditions. As expected, crash types related to high speeds, such as “Run Off” and “Hitting Animals” were less common for congested conditions. Crashes coded as “Other” were higher for oversaturated conditions. In the database, those encompassed: fires/explosions, flooding, landslides and vehicles being hit by objects/debris. These crash subtypes have a strong component of randomness so that, the

more vehicles travelling on the road, the more likely it was that an unexpected adverse event happened to one of them.

Figure 5.3 – Distribution of single-vehicle crashes for LOS A-E vs LOS F.

5.2 MODELLING THE RELATIONSHIP BETWEEN DENSITY AND CRASH RATE

This section focuses on the development of a model to predict crash rate as a function of traffic density and other influencing variables. The first step consists of a statistical analysis on the variables available in the database to determine which ones should be included in the model, and at what levels or categories. The second step is focused on building regression models using the selected variables. The model developed is examined for consistency and accuracy relative to the traditional HSM methodology. Finally, a discussion on the implications of the findings is provided.

5.2.1 INFLUENCING FACTORS

The principle of the model building process proposed is to find a balance between having as many policy-sensitive explanatory variables as possible and keeping the model consistent and functional. The process used to seek this balance is described as follows.

The data were subdivided into 3 pc/km/lane traffic density bins according to the levels or categories of the candidate variables to be added to the model, so that each bin reflects a unique combination of density range and other characteristics. For each bin, the number of crashes,

exposure and crash rates were computed as shown in Equation 5.2. If variables and levels of variation are added to the model, more combinations of characteristics are created. This increases the number of individual data bins, at the expense of reducing the sample for each. The required sample n (number of observed crashes for one bin) within each density range for reliable modeling purposes would be:

$$n = \left[\frac{(Z_{\alpha/2})\sigma}{E} \right]^2 \quad 5.3$$

where $Z_{0.025}$ for a desired confidence of $C = 1 - 0.05$ is 1.96; σ is the standard deviation of the crash rates within each density range; and E is an accepted margin of error, established to be 1 crash per million veh-km in this research.

To select which variables should be considered for inclusion in the model, the following procedure was followed, for each variable:

- a) Two-Way ANOVA analysis was performed to compare each variable (treatment) and their possible interaction with traffic density (blocks);
- b) Pairwise comparisons using Tukey's statistic were made to compare all categories within each variable. When two categories could not be considered statistically different (could not reject H_0) at 95% confidence, they were grouped, and the ANOVA test was conducted again;
- c) If the H_0 could not be rejected for the minimum number of categories (2), or if grouping was not reasonable, no conclusions could be reached regarding the effect of that variable on the crash rate-density relationship.

Table 5.1 shows the ANOVA results, as well as the means and 95% confidence intervals of each category and variable that remained in the model: Segment Type, Area Type, Posted Speed and Nighttime. The differences between those categories can be also observed in Figure 5.4.

Among the tested factors, there was not sufficient evidence to support that the number of lanes, percent of trucks or weekend traffic affect the studied relationships.

Regarding the segment type, only weaves could be considered different from other types of segments at 95% confidence. The relationship for weaving segments, although somewhat

fuzzy, also follows the U-shape and has a higher crash rate as compared to other segment types, especially for high density conditions.

Overall, urban sites could not be considered different from rural sites ($F_{\text{obs}} = 4.07 < F_{0.05,1,10} = 4.96$). However, an interaction to density was observed ($F_{\text{obs}} = 23.73 < F_{0.05,10,1118} = 1.84$), noticeably for low traffic density ranges. For that reason, this variable was considered for inclusion in the model.

Table 5.1 – Influencing variables 2-Way ANOVA statistics.

Variable	Category	Mean	W _{Tukey}	95% Confidence Interval		
				LB	UB	
Segment Type	Expressway Segment	1.00	0.97	0.03	1.97	
	Weavings	2.68		1.71	3.65	
Area Type	Rural	0.92	0.26	0.66	1.18	
	Urban	1.19		0.93	1.45	
Posted Speed (km/h)	80-90	1.60	0.38	1.22	1.98	
	100-120	0.99		0.61	1.37	
Nighttime	No	0.94	0.15	0.79	1.10	
	Yes	1.16		1.00	1.31	
Source	Df	SS	MS	F	F _{crit}	p-value
Segment Type	1	854.61	854.61	11.62	4.96	0.007
Density	10	9669.71	966.97	13.14	2.98	0.000
Interaction	10	735.72	73.57	886.30	1.84	0.000
Error	1188	98.62	0.08			
Total	1209	11358.65				
Area Type	1	21.48	21.48	4.07	4.96	0.071
Density	10	1932.06	193.21	36.57	2.98	0.000
Interaction	10	52.82	5.28	23.73	1.84	0.000
Error	1188	264.42	0.22			
Total	1209	2270.79				
Posted Speed	1	111.29	111.29	9.70	4.96	0.011
Density	10	4218.50	421.85	36.78	2.98	0.000
Interaction	10	114.69	11.47	50.58	1.84	0.000
Error	1188	269.36	0.23			
Total	1209	4713.84				
Nighttime	1	14.12	14.12	7.47	4.96	0.021
Density	10	2006.53	200.65	106.10	2.98	0.000
Interaction	10	18.91	1.89	8.05	1.84	0.000
Error	1188	279.08	0.23			
Total	1209	2318.64				

Figure 5.4 – Relationships between density and crash rate, by influencing variable.

Concerning posted speeds, after the process described in this subsection, two functional groups could be identified. The 100-120 km/h segments comprise mostly rural expressways with

modern design, or urban freeways segregated from the local road system and only accessible through interchanges spaced at long distances. The 80-90 km/h segments encompass urban multilane highway segments. For these, an overall higher crash rate was observed for all traffic densities, so that H_0 could be rejected, and the two classes could be considered statistically different at 95% confidence.

Nighttime conditions were found to be related to slightly higher crash rates for all density intervals.

5.2.2 PROPOSED MODEL

Regression analysis was used next to build model alternatives, based on the candidate variables described in the previous subsection. Each observation of the model corresponded to a data bin for which the sample size was considered statistically significant according to Equation 5.3. To use polynomial relationship between crash rates and density in a linear regression method, two transformed variables were created: cubic and quadratic densities (Density^3 and Density^2 , respectively). Interaction variables were also tested, by relating the original variables to density variables. The criteria used to select the proposed model were: (a) F statistics as compared to the simplified model; (b) all model variables must be different from zero and statistically significant, at 95% confidence (t value > 1.96), with the possible exception of the model constant; c) low collinearity with other variables in the model.

The proposed regression equation ($F_{7,62} = 34.089$, $p < .000$) has an adjusted R^2 value of 0.763. Table 5.2 shows the model coefficients and ANOVA statistics for the selected model. Apart from traffic density, weaves, lower-standard urban multilane highways (posted speed 80-90 km/h) and nighttime conditions were found to increase crash rates. By adding these explanatory variables to the model, much of the asymmetry between low and high density ranges that was described by the cubic term of the equation in previous work (Figure 2.1) could be explained by the interaction variables (ex.: $\text{Density}^2 \times \text{Urban}$).

Table 5.2 – Model coefficients and ANOVA statistics.

Variable	B	Std. Error	t	Sig.
(Constant)	.25504	.134	1.908	.061
Density ²	.00179	.000	4.321	.000
Density ³ X Weaving	.00023	.000	10.599	.000
Urban	1.87078	.206	9.101	.000
Density X Urban	-.22427	.024	-9.340	.000
Density ² X Urban	.00538	.001	6.402	.000
Density X Posted Speed 80-90 km/h	.02701	.007	3.756	.000
Nighttime	.31655	.095	3.338	.001

ANOVA					
Source	df	SS	MS	F	p-value
Regression	7	36.729	5.247	34.089	0.000
Error	65	10.005	.154		
Total	72	46.733			

The proposed model to predict total crash rate, in 1 million veh-km, is as follows:

$$\begin{aligned} \text{Crash Rate} = & 0.25504 + 0.00179D^2 + .00023D^3W + 1.87078U - 0.22427DU \\ & + 0.00538D^2U + 0.02701DPS + 0.31655N \end{aligned} \quad 5.4$$

where D is the traffic density; $W = 1$ when the segment is a weaving and 0 otherwise; and $PS = 1$ when the segment is an urban multilane highway (posted speed 80-90 km/h), and 0 otherwise; (urban freeways or rural expressways with posted speed higher than 90 km/h); $U = 1$ when abutting area is urban and 0 otherwise; and $N = 1$ for nighttime conditions and 0 for daytime.

Finally, Figure 5.5 shows the analysis of the standardized residuals of the model. The Kolmogorov-Smirnov test was performed to check the assumption that the residuals follows a normal distribution. The test statistic obtained was 0.096, with a p-value of 0.095, indicating that H_0 cannot be reject, and that the residuals follow roughly a normal distribution, with a few values over 2.5 [Figure 5.5 (a)]. The absence of evident biases can be confirmed in Figure 5.5 (b), in which no trend could be identified.

Figure 5.5 – Analysis of standardized residuals.

Figure 5.6 presents the application of the proposed model, highlighting the effect of some characteristics to the predicted number crash rates. Figure 5.6 (a) shows that, unlike previous findings, rural expressway segments don't follow a U shape format, suggesting that low-volume rural roads operate safely in most cases. Weaves and urban highways, on the other hand, were associated with higher crash risk for both low and high density conditions, reaching minimum crash rates at LOS B-C [Figure 5.6 (b)]. For weaves, safety conditions rapidly deteriorate after this threshold. This difference may be associated with less forgiving roadway and roadside designs to accommodate higher speeds that are typical at low density conditions, and can be better handled on the rural freeway environment.

In urban areas, the upgrade from a multilane to a freeway design could improve safety for all density levels. Finally, nighttime is associated with a slightly increased crash risk.

Figure 5.6 – Application of the proposed model.

(a) Rural Expressways vs Urban Multilane Highways

(b) Urban Freeways vs Weaves

5.3 COMPARISON BETWEEN THE PROPOSED MODEL AND THE HSM

By multiplying the crash rate obtained by Equation 5.4 by the traffic volume in each interval and segment length, the predicted number of crashes is calculated. The sum of the predicted number of crashes for all time intervals within a year yields the annual predicted number of crashes, which is the same output produced by the Highway Safety Manual (HSM) methods.

In order to compare the proposed equation to the HSM, the annual number of crashes for each segment was calculated using both methods and compared to field values. AADT values ranged from 3,680 to 93,070 veh/day (average of 30,980). The HSM method for multilane highways (AASHTO, 2001) was used for all segments. The model was calibrated for the studied region, yielding an adjustment factor $Cr = 2.1$.

Figure 5-7 presents the results of this analysis. Figure 5-7 (a) compares the annual number of crashes estimated by the calibrated HSM and the equation developed in this research, while Figure 5-7(b) shows the distribution of errors relative to the field data for both methods. Although the proposed model produces estimates that deviate somewhat from the calibrated HSM to an extent, the distribution of errors follows a similar pattern.

Figure 5.7 – Comparison of results from the proposed equation and the HSM.

These results suggest that the proposed equation is capable of producing estimates on an hourly basis that are still consistent when computed annual estimates.

5.4 DISCUSSION

The findings of this work and the proposed model have applications for various types of safety and mobility studies. Differently from planning level methods, in which the number of crashes is predicted in an annual basis, this work aims to provide relationships to hourly traffic densities, thus allowing for the assessment of crash risk.

From the design perspective, minimum crash rates are associated to LOS A for rural expressways, and B to C for urban expressways. Also, the risks associated with weaving segments and the benefits from the improvement of an urban multilane highway to a freeway can be estimated.

At the operations level, the proposed model can be of use in applications related to ATM. Strategies such as ramp metering could be adjusted to find best balances between delays and crash risk. This is particularly true when weaving segments exist along the mainline freeway.

For planning studies using the HCM (TRB, 2016), Equation 5.4 could be used to estimate crash rates to be used in the generation of scenarios for reliability analysis, still accounting for factors such as are type and day and night conditions.

5.5 FINAL REMARKS

This research developed a novel model to describe the relationship between hourly traffic density and crash rates on expressways, composed of rural and urban multilane highways, basic freeway segments, ramp influence areas and weaving segments, based on a database containing over 35 million traffic observations and more than 20,000 crashes. Preliminary analysis confirmed that the curve for the entire database follows a U shape, which is similar to previous studies. This shape is the consequence of a convex downward and a convex upward curve for single- and multi-vehicle crashes, respectively.

Regarding crash severity, the proportion of fatal and injury crashes (FI) and PDO crashes were similar for LOS A to C, reaching a minimum between LOS B and C. For congested conditions, PDO crashes are more prevalent.

The analysis of influencing factors revealed that weaves, urban areas, nighttime conditions and urban multilane highways (segments with 80-90 km/h posted speed) are related to higher crash rates, at 95% confidence. Those factors also explained much of the asymmetry of the curves, described by interaction variables.

Using the selected variables, a regression equation was fitted. The application of the proposed model showed that, unlike urban expressways, the density-crash rate relationship for rural expressway segments doesn't follow a U shape format, suggesting that low volume rural roads are safer than the higher volume ones. In urban areas, the upgrade from a multilane to a freeway design could improve safety for all density levels.

The comparison with the HSM and field data suggests that the proposed equation is capable of producing estimates on an hourly basis that are still consistent when computed year-round projections.

The findings of this work have implications to policy and design decisions, and the produced equation could be incorporated to active traffic management (ATM) and HCM reliability analysis.

6 ESTIMATION OF CRASH SEVERITY ON EXPRESSWAYS

Although studies on crash severity concentrate in developed countries, the majority of deaths occur in low and middle income countries (Malveira et al., 2015; Çelik and Oktay, 2017), specially on highways and rural roads, even while carrying less traffic than urban streets (ROSPA, 2017). The severity of a crash defines the consequence of a traffic incident, ranging from property-damage to fatalities in the worst case. In order to understand the factors that affect crash severity and to design countermeasures to prevent injuries and fatalities, researchers and engineers have been using of multivariate statistics to account for the interactions between the several variables that might play a role in the gravity of a road crash.

While many studies focused on geographic, socioeconomic and other human factors as the main cause of crashes (Quddus, 2015), more recent research recognize the interactions between more factors, including policy sensitive roadway and environmental aspects that can be accounted for when designing safety countermeasures. However, only Quddus, Wang and Ison (2010) relied on a database that could combine prevailing traffic conditions to other roadway and environmental data, for one facility in the UK.

This work was motivated by the absence of stablished models capable of predicting severity distributions simultaneously as a function of prevailing traffic conditions and other explanatory variables. The achievement of such goal was made possible by the availability of the database described in Chapter 4. This will allow for better quantification of the impacts of safety countermeasures and other policy-sensitive variables to crash severities and frequencies, which could support transportation feasibility and cost-benefit studies, warranting investment strategies based not only in mobility, but also safety criteria.

In this context, the specific objectives of this stage of the research are to: (a) develop crash severity distribution models for expressways as a function of prevailing traffic conditions and other explanatory variables, using a modeling framework adequate to the crash severity analysis problem; (b) compare the results to a generic crash severity distribution model, in order to verify the consistency of the proposed model; (c) discuss the marginal impact of the significant variables and the implications for safety analysis practices.

6.1 METHODOLOGY OVERVIEW

In this work, the Ordered Multinomial Logit Model is used to investigate the relationship between the crash severity probabilities and the explanatory variables available in the database. As indicated in the literature, this model is typically applied to problems in which the outcome is discrete and ordered in a specific sequence, in the case, from fatal (F), to injury (I), to property-damage only (PDO) crashes.

The concept of utility used in other types of choice models is replaced by an index function of “propensity”. Given a set of ordered outcome alternatives c_q , a set of thresholds μ_n is defined equal to the number of alternatives k minus 1:

$$c_q = \{0, 1, 2, \dots, k - 1\} \quad 6.1$$

The thresholds in this case are such that:

$$-\infty < \mu_F \leq \mu_I \leq \mu_{PDO} < \infty \quad 6.2$$

The probability that an injury crash happens is given by:

$$P_q(B) = \Lambda(\mu_F - V_q) - \Lambda(\mu_{PDO} - V_q) \quad 6.3$$

where V_q is the propensity function given the attributes q . To define the propensity functions, the process described above was followed.

First, the database was cleaned, checked for consistency and divided in three subsets: (a) SV crashes; and (b) MV crashes.

Only variables that could possibly be associated with each subset were used. For example, crash types head-on, sideswipe, angle or rear-end are only possible for events involving more than one motorized vehicle and are therefore discarded from the analysis of single-vehicle crashes. The models were fitted using variables that describe prevailing traffic conditions, such as density and average speed, not individual vehicles types. The intention was to develop models that could be used to generate severity distribution functions based on characteristics that can be observed in the field.

The statistical analysis itself was made separately for SV and MV crashes, using an ordinal regression technique. In the first step, all explanatory variables in the database were considered. A 95% confidence was used to determine the statistically significant variables ($p\text{-value} < 0.05$). In each subsequent step of the procedure, one variable with the highest p -value was eliminated, until all remaining variables could be considered statistically significant. The statistical significance of the model as a whole and of each thresholds μ was also verified in each step.

6.2 RESULTS

Table 6.1 shows the results of the statistical analysis, for SV and MV crashes. The table lists all variables considered statistically significant for at least one of the subsets. The coefficients and statistics are described only for the variables that remained in the final models, for each type of crash. The table also shows the resulting Log-likelihood of the models. The three severity categories are divided by two μ thresholds (Equation 6.1).

Runoffs (SV) and head-on (MV) were considered the most severe types of crashes. Lower visibility conditions, reflected by nighttime and rain affected all models. Regarding the geometry aspects, weaves and bridges were related to crash severity for MV crashes, while the presence of work zones, guardrails and barriers affected the outcome of SV crashes, which is consistent with runoffs being the gravest type of single-vehicle crash.

The assessment of the influence of prevailing traffic conditions revealed that while the severity of MV crashes is linked to the traffic density and the variability of density through time, the severity of SV crashes is more closely related to average speed.

Geometry variables that were analyzed but not found to be statistically significant were the presence of upgrades/downgrades. This may be due to two factors: (a) there are not enough crash data in these segments to draw statistically significant conclusions; (b) the studied expressways have high design standards, and thus these segments provide the users practically the same level of safety and comfort as the remaining segments. Regarding environment variables, the no association was made between the “visibility” variable and crash severity. This may be due to imprecisions regarding visibility classification in the field by crash victims and officers.

Table 6.1 – Results and significant variables for each model.

Variable	Single-vehicle crashes				Multiple-vehicle crashes			
	B	Std. Error	Wald	p-value	B	Std. Error	Wald	p-value
Crash Type								
Collision with object	-1.948	.150	168.004	.000				
Hit animal	-2.952	.215	187.899	.000				
Runoff	-1.098	.090	150.145	.000				
Head-on					1.171	.194	36.583	.000
Rear-end					-.228	.067	11.469	.001
Sideswipe					-.208	.076	7.566	.006
Time and Environment								
Year	.041	.018	5.338	.021				
Weekend					.211	.057	13.778	.000
Night	-.182	.064	7.974	.005	.211	.051	16.974	.000
Drizzle	-.401	.198	4.101	.043				
Rain	-.868	.093	86.439	.000	-.438	.090	23.632	.000
Urban area	.173	.073	5.566	.018	.257	.052	24.692	.000
Geometry and design								
Number of Lanes	.122	.043	8.252	.004				
Sharp Curve	-.251	.123	4.186	.041	-.788	.148	28.291	.000
Weave					.607	.157	14.845	.000
Bridge segment					1.259	.553	5.179	.023
Guardrails	-.708	.077	84.823	.000				
Work zone	-.345	.083	17.286	.000				
Traffic Variables								
Traffic density					-.030	.004	59.505	.000
Traffic density variation					.009	.004	4.174	.041
Average speed	-.007	.003	7.273	.007				
Number of vehicles					.144	.033	18.685	.000
Percent of Trucks	-.006	.002	8.913	.006				
μ_1 (Injury threshold)	-1.737	.306	32.164	.000	.519	.105	24.416	.000
μ_2 (Fatal threshold)	2.908	.326	79.505	.000	4.173	.135	953.854	.000
Model Fitting Information								
N		4,782				7,648		
-2 Log-likelihood at convergence		6392.373				10599.182		
-2 Log-likelihood at equal shares		7158.493				10933.718		
Chi-Square		766.119				334.536		
p-value		0.000				0.000		

6.3 MARGINAL EFFECTS

The magnitude of the coefficients B shown in Table 6.1 provides a preliminary indication on the importance of each variable in the models. However, contrary to traditional linear regression modeling, neither the magnitude nor the sign of the variables can be directly used to make inferences on the marginal effect of each variable to the discrete choice model response (Greene and Hensher, 2009).

Table 6.2 shows the marginal effects of each variable on the results. The “base” scenario reflects the average probability of each severity level obtained by applying the models proposed in this study to all crashes in the database. Each row, labelled after one variable tested, represent one scenario created by changing only that variable while keeping all other variables constant. For example, the row “night” would represent the average probabilities of each severity level in nighttime conditions. By comparing the altered probabilities to the ones from the base scenario, it is possible to assess the impact of each variable on crash severity.

Table 6.2 – Probabilities estimated for each non-continuous variable.

Variable	Single-vehicle crashes			Multiple-vehicle crashes		
	PDO	Injury	Fatal	PDO	Injury	Fatal
Base	52.4%	46.4%	1.2%	60.8%	37.4%	1.8%
Crash Type						
Collision with object	70.2%	29.3%	0.3%			
Hit animal	86.4%	13.4%	0.0%			
Runoff	56.7%	42.5%	0.8%			
Head-on				31.1%	62.4%	6.5%
Rear-end				62.2%	36.2%	1.6%
Sideswipe				62.9%	35.5%	1.6%
Time and Environment						
Weekend				53.8%	43.8%	2.4%
Night	56.4%	42.5%	1.0%	55.4%	42.4%	2.2%
Drizzle	62.2%	37.2%	0.7%			
Rain	69.5%	30.0%	0.3%	69.5%	29.3%	1.2%
Urban area	47.8%	50.7%	1.5%	59.4%	38.7%	1.9%
Geometry and design						
Sharp Curve	52.9%	45.9%	1.2%	76.0%	23.1%	1.0%
Weave				47.4%	49.6%	3.0%
Bridge segment				28.2%	64.1%	7.6%
Guardrails	66.1%	33.4%	0.5%			
Work zone	54.4%	44.5%	1.1%			

Figure 6.1 graphically presents this information, making it possible to visually compare the different severity levels. The further the lines are from the center, the higher the probability of that severity level as compared to the other levels. The following subsections discuss each variable group in more detail.

Figure 6.1 – Comparison of results from the proposed equation and the HSM.

6.3.1 CRASH TYPE

As mentioned before, different crash types were associated to the severity of single-vehicle or multiple-vehicle crashes. In no model collisions with objects in the road were considered to be statistically different from any other type of crash. This may be due to the varied nature of objects that can be possibly hit, from large animals to stationary road work equipment. Runoffs,

on the other hand, proved to be a more severe type of crash, especially for single-vehicle crashes. For multiple-vehicle crashes, angle and head-on crashes were considered more severe than other accidents, at 95% confidence. Although head-on accidents in multilane highways are not common, they are a possibility when there is no physical barrier on the median, and significantly more severe than the average, according to the analysis. Finally, sideswipe and rear-end accidents could not be considered any different than other types of crashes.

6.3.2 TIME AND ENVIRONMENT

Weekend was considered statistically significant for the MV model, slightly increasing the probability of more severe accidents; similarly, crashes in urban areas were found to be more severe for both SV and MV crashes. Although the cause of this is unclear, it might be related to substance abuse or conflicts between highway and urban traffic. While nighttime increased crash severity for MV crashes, this condition was associated to less severe crashes for the SV model. Similarly, rainy conditions were linked to less severe crashes, which could be explained by the extra caution that drivers use under low visibility conditions.

6.3.3 GEOMETRY AND DESIGN

With respect to geometry, it was not possible to affirm that grades affect the severity of accidents with statistical confidence. On the other hand, curved segments were associated with less severe MV crashes. Note that these segments are usually part of expressways on rolling terrain with consistent design standards, so that lower severities could be explained by lower speeds and extra attention from the drivers' part. Weaves and bridge segments were found to significantly increase crash severity for MV crashes, while the guardrails and barriers proved effective in reducing the probability of severe SV crashes. This finding reinforces the role of guardrails and other types of physical barriers as a countermeasure of single-vehicle, runoff accidents. Although these cannot eliminate crashes, a well-designed device may potentially reduce the consequences of the accident, from FI to PDO.

6.3.4 TRAFFIC CHARACTERISTICS

Figure 6.2 presents the relationships between traffic density and crash severity for SV and MV crashes. Each small dot represents the estimate for one observation in the crash database, while the “base” values is the average severity distribution for the whole database, for each density interval of 3 pc/km/lane, obtained from the models calibrated in Chapter 5 (Figure 5.1).

Figure 6.2 – Relationships between traffic density and crash severity distribution.

The variability of the results reflects the effect of the multiple factors accounted for by the fitted models. It can be noticed that, for MV crashes, there is strong evidence of the relationship between traffic density and crash severity, evidenced by the lower deviation from the average values. The relationship between density and severity of SV crashes can be roughly observed until the range between 5-10 pc/km/h, after which results become much more dependent on other explanatory variables, such as the area type.

Figure 6.3 shows the relationships between average travel speed and severity distribution. In this figure, fatal and injury crashes are grouped as FI crashes. In the range of 80-100 km/h, the distribution of PDO and FI crashes is more homogeneous. For both lower and higher speed ranges, the variability is greater, with a higher proportion of PDO crashes for high speed. This phenomenon may be explained by the fact that higher speeds are possible in higher design standards expressways. Conversely, lower speeds are more frequent on more heterogeneous highways, including urban segments.

Figure 6.3 – Variability of the relationship between speed and SV crash severity.

Figure 6.4 illustrates the variability of the crash severity distributions for different values of density difference, defined as the variation, in pc/km/lane, between the density right prior the crash and the moment immediately before it. The proportion of PDO and FI crashes is the most balanced when the density variation is lower, near zero. Negative density differences (left side of the graph) reflect discharge conditions, resulting in proportionally less FI crashes as compared to PDO crashes. Positive values of density variation (right side of the graph) represents congestion formation. For this situation, the variability of crash severity distributions increase and are more difficult to predict.

Figure 6.4 – Relationship between density differences and MV crash severity.

6.4 ALTERNATIVE MODEL

Following the practice from many previous studies, two alternative models were created, in which the specific vehicle types involved in the crash are taken as explanatory variables. In this study, however, the variables that describe prevailing traffic conditions are maintained to study the possible interactions between the surrounding traffic and vehicles involved in the crash. Table 6.3 shows the results for these models the MV and P&B.

Table 6.3 – Results and significant variables – Alternative MV and P&B Models.

Variable	Multiple-vehicle crashes				Pedestrian & Bicycles			
	B	Std. Error	Wald	p-value	B	Std. Error	Wald	p-value
Crash Type								
Head-on	1.197	.197	36.965	.000				
Rear-end	-.188	.071	7.037	.008				
Sideswipe	-.173	.080	4.634	.031				
Time and Environment								
Year	.054	.013	18.315	.000				
Weekend	.251	.060	17.380	.000				
Night	.225	.054	17.098	.000	.982	.256	14.675	.000
Rain	-.368	.094	15.269	.000				
Geometry and design								
Sharp Curve	-.654	.161	16.428	.000				
Weave	.469	.164	8.242	.004				
Bridge segment	1.239	.561	4.886	.027				
Traffic Variables								
Traffic density	-.046	.004	106.631	.000				
Average speed	-.012	.002	47.176	.000				
Number of vehicles	.152	.035	18.847	.000				
Involved Vehicles								
Auto	-.559	.053	112.488	.000				
Bus	-.311	.144	4.661	.031				
Truck	-.235	.063	13.774	.000				
Motorcycle	2.411	.096	632.009	.000	-1.176	.538	4.778	.029
μ_1 (Injury threshold)	-.680	.219	9.608	.002	-4.029	.528	58.220	.000
μ_2 (Fatal threshold)	3.416	.236	208.638	.000	1.090	.223	23.933	.000
Model Fitting Information								
N		7,648				343		
-2 Log-likelihood at convergence		10004.808				24.031		
-2 Log-likelihood at equal shares		11328.486				45.826		
Chi-Square		1323.678				21.795		
p-value		0.000				0.000		

The parameters for MV crashes are similar to the ones previously observed, though part of the phenomenon is explained according to the vehicles involved in the crash, especially when for crashes involving at least one motorcycle. For the P&B model, the analysis of the data was not sufficient to draw broader conclusions, but indicated that crashes involving motorcycles tend to be less severe, and crashes at night, more severe.

6.5 FINAL REMARKS

This chapter modeled the impact of a series of explanatory variables to predict crash severity, using the ordered response choice model and the database described in Chapter 4. The framework that better fit this database led to the development of two different models: single-vehicle crashes (SV) and multiple-vehicle crashes (MV). The analysis of the results showed that the factors that explain the severity of crashes varies widely between these models, as the causes for SV and MV crashes are different.

Head-on crashes were found to be the most severe crash type for the MV model, while run-offs were more significant for the SV model. Consequently, the use of guardrails and barriers as a safety countermeasure proved to effectively reduce severity for SV crashes. Other design features found to be statistically significant were weave and bridge segments, which were associated with more severe MV crashes. Finally, conditions that impair visibility, such as rain and nighttime were linked to more severe MV crashes, but slightly less severe SV crashes.

The unique database used in this study allowed the investigation of the influence of prevailing traffic conditions on crash severity, while controlling for other factors. The results suggested that MV crash severity is negatively related with traffic density, while SV crashes are more closely related to speed. The variability of the crash severity estimates for low speed and high speed ranges proved to be high, reflecting the impact of other explanatory variables on low and high speed roads.

For future studies, it recommended that the results obtained in this research are translated into crash severity functions that can be integrated in crash prediction models which are able to predict crash frequency as a function of roadway and environmental conditions, as well as prevailing traffic conditions.

7 CONCLUSIONS AND RECOMMENDATIONS

The goal of this research was to model the relationship between crash rates and different operating conditions for two different components of a road network: corridors with signalized intersections and expressway (uninterrupted) segments.

The first part of this project produced a new methodology capable of combining mobility, safety and emissions measures to the objective function of a genetic algorithm now capable of optimizing signal timing and coordination on arterials. The resulting methodology is available to the users as part of the HCS – Streets module. The selected safety model has the advantage of considering geometry and control aspects, while its multiplicative structure allows for local calibration of existing parameters and future inclusion of the effects of additional factors, such as HSM Crash Modification Factors, the number of phases, maximum greens or the presence of countdown timers, in example.

It was also presented the results from a sensitivity analysis conducted to provide insight on the effects and order of relevance of 20 key variables on the model's outcomes and the associated trade-offs between mobility, safety, and emission. This insight will help the designer, signal control engineer, and traffic analyst when designing intersection geometry and signal control.

The subsequent parts of this research project were part of an effort to develop innovative models able to predict crash frequencies and severities as a function of prevailing traffic conditions and other explanatory variables. This goal was only possible due to the availability of a unique database of Brazilian expressways, composed of rural and urban multilane highways, basic freeway segments, ramp influence areas and weaving segments, with over 35 million traffic observations and more than 20,000 crashes.

In the second part of the research, it was developed a novel model to describe the relationship between hourly traffic density and crash rates Preliminary analysis confirmed that the curve for the entire database follows a U shape, which is similar to previous studies. This shape is the consequence of a convex downward and a convex upward curve for single- and multi-vehicle crashes, respectively.

Further analysis of the influencing factors revealed that weaves, urban areas, nighttime conditions and urban multilane highways (segments with 80-90 km/h posted speed) are related

to higher crash rates, at 95% confidence. Using the selected variables, a regression equation was fitted. The application of the proposed model showed that, unlike urban expressways, the density-crash rate relationship for rural expressway segments doesn't follow a U shape format, suggesting that low volume rural roads are safer than the higher volume ones. In urban areas, the upgrade from a multilane to a freeway design could improve safety for all density levels.

In the last part of the research, it was modeled the impact of a series of explanatory variables to predict crash severity, using the ordered response choice model the database described in Chapter 4. The framework that better fit this database led to the development of two different models: single-vehicle crashes (SV) and multiple-vehicle crashes (MV). The analysis of the results showed that the factors that explain the severity of crashes varies widely between these models, as the causes for SV and MV crashes are different.

Head-on crashes were found to be the most severe crash type for the MV model, while run-offs were more significant for the SV model. As a consequence, the use of guardrails and barriers as a safety countermeasure proved to effectively reduce severity for SV crashes.

The investigation of the influence of prevailing traffic conditions on crash severity revealed that multiple-vehicle crash severity is positively related with traffic density, while single-vehicle crashes are more closely related to speed. Still, the variability of the crash severity estimates for low speed and high speed ranges proved to be high, reflecting the impact of other explanatory variables on low and high speed roads.

The findings of this work have implications to policy and design decisions, and the produced equation could be incorporated to active traffic management (ATM) and HCM reliability analysis, still accounting for factors such as are type and day and night conditions.

For future work, it is recommended that the research done on freeway systems and multilane highways to be implemented and tested in a tool, as done for the methodology developed for signalized arterials. This could allow users to benefit from the findings of this research, as well as give insight of the impacts of the variables considered in the modelling to transportation planning and operations.

Also, the models developed for expressways could potentially be extended to other types of facilities, such as two-lane highways. This would allow the assessment of the ideal traffic and

design standards that justify the improvement of a two-lane highway to a multilane highway. Also, other variables that could be incorporated to the crash prediction and crash severity models would require more detailed analysis on the roadway design, and could include median types and widths, rumble strips, ramp/interchange density and detailed horizontal and vertical alignment information, such as the precise curve radii and grade magnitudes and lengths along all studied network.

Finally, it is recommended that the relationships for pedestrian and bicycles to be further investigated, given the high fatality rate of these crash types. This will require not only motorized vehicle data, but also non-motorized traffic on curbsides and crossings.

8 REFERENCES

- Abdel-Aty, M. (2003) Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research*, 34, 597-603.
- Agbelie, B., and Roshandeh, A. (2015) Impacts of Signal-Related Characteristics on Crash Frequency at Urban Signalized Intersections. *Journal of Transportation Safety & Security*, 7, 199-207.
- AASHTO (2010) Highway Safety Manual. American Association of State Highway and Transportation Officials, Washington, D.C.
- Abdel-Aty, M. (2003) Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research*, 34, 597-603.
- Archer, J. (2004) Methods for the Assessment and Prediction of Traffic Safety at Urban Intersections and their Application in Micro-Simulation Modeling. PhD dissertation, Division of Transport and Logistics, Royal Institute of Technology, Stockholm, Sweden.
- Arona, M., Billot, R., el Faouzi, N. E., and Seidowsky, R. (2015) Traffic indicators, accidents and rain: some relationships calibrated on a French urban motorway network. *Transportation Research Procedia*, 10, 31-40.
- Azar, S., Reynolds, B., and Narayanan, S. (1999) Comparison of Two Objective Optimization Techniques with and within Genetic Algorithm. *Proceedings of the ASME Design Engineering Technical Conferences*.
- Beasley, D., Bull, D., and Martin, R. (1993) An Overview of Genetic Algorithms: Part 1, Fundamentals. *University Computing*, 15 (2), 58-69.
- Campolongo, F., Cariboni, J., and Saltelli, A. (2007) An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10), 1509-1518.
- Çelik, A. K., and Oktay, E. (2017) A comparison of ordered and unordered response models for analyzing road traffic injury severities in the North-Eastern Turkey. *Periodica Polytechnica Transportation Engineering*, 45 (3), 119-132.
- Chang, J., Oh, C., Chang, M. (2000) Effects of Traffic Condition (v/c) on Safety at Freeway Facility Sections. *Transportation Research Circular E-C018*, 200-208.
- Chen, X., Hadi, M., Xiao, Y., and Elefteriadou, L. (2016) Development of Macroscopic Emission Estimation Model Based on Microscopic Operating Modes. *Transportation*

Research Record, 2570, 39-47.

Chin, H., and Quddus, M. (2003) Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis and Prevention*, 35, 253–259.

Coutinho, C. H. L., Cunto, F. J. C.; and Ferreira, S. (2015) Análise da severidade dos acidentes com motocicletas utilizando modelos probit e logit ordenados. *Transportes*, 23, 60-66.

Daniel, J. R., and Maina, E. (2011) Relating Safety and Capacity on Urban Freeways. *Procedia Social and Behavioral Sciences*, 16, 317-328.

Essa, M., and Sayed, T. (2015) Simulated Traffic Conflicts. Do They Accurately Represent Field-Measured Conflicts? *Transportation Research Record*, 2514, 48-57.

Frantzeskakis, J. M., and Lordanis, D. I. (1981) Volume-to-Capacity Ratio and Traffic Accidents on Interurban Four-Lane Highways in Greece. *Transportation Research Record*, 1112, 29-38.

Geedipally, S. R., Bonneson, J. A., Pratt, M. P., and Lord, D. (2013) Severity Distribution Functions for Freeway Segments. *Transportation Research Record*, 2398, 19–27.

Gettman, D., Pu, L., Sayed, T., and Shelby, S. (2008) Surrogate Safety Assessment Model and Validation. Final report. FHWA-HRT-08-051. FHWA, U.S. Department of Transportation.

Golob, T. F., and Recker, W. W. (2003) Relationships Among Urban Freeway Accidents, Traffic Flow, Weather, and Lighting Conditions. *Journal of Transportation Engineering*, 129 (4), 342-353.

Greene, W. H., and Hensher, D. A. (2009) Modeling Ordered Choices: A Primer. DOI: 10.1017/CBO9780511845062.

Hadi, M., Elefteriadou, L., Chen, X., Wang, T., and Xiao, Y. (2017) Signal Optimization with Consideration of Environmental and Safety Impacts, Part A. STRIDE Final Report, Gainesville, FL.

Harwood, D. W., Bauer, K. M., and Potts, I. B. (2013) Development of Relationships Between Safety and Congestion for Urban Freeways. *Transportation Research Record*, 2398, 28-36.

Instituto de Astronomia, Geofísica e Ciências Atmosféricas (2018) Nascer e ocaso do Sol

- (2011-2025). University of São Paulo. Retrieved on July 2018 from <http://www.iag.usp.br/astrologia/nascer-e-ocaso-do-sol>.
- Islam, M. R., Wyman, A. A., and Hurwitz, D. S. (2017) Safer driver responses at intersections with green signal countdown timers. *Transportation Research Part F* 51, 1-13.
- Kockelman, K. M., and Kweon, Y. J. (2002) Driver injury severity: An application of ordered probit models. *Accident Analysis and Prevention*, 34 (3), 313-321.
- Kononov, J., Lyon, C., Allery, B. K. (2011) Relating Flow, Speed and Density of Urban Freeways to Functional Form of an SPF. 90th Transportation Research Board Annual Meeting, Washington, D.C.
- Kuehl, R. (1999) *Design of Experiments: Statistical Principles of Research Design and Analysis* 2nd ed. Duxbury Press.
- Lambert, D. (1992) Zero-Inflated Poisson Regression with an Application to Defects Manufacturing. *Technometrics*, 34 (1), 1-14.
- Lee, C., and Li, X. (2015) Predicting Driver Injury Severity in Single-Vehicle and Two-Vehicle Crashes with Boosted Regression Trees. *Transportation Research Record*, 2514, 138–148.
- Lord, D., Manar, A., Vizioli, A. (2005) Modeling crash-flow-density and crash-flow-VC ratio relationships for rural and urban freeway segments. *Accident Analysis & Prevention*, 37, 185-199.
- Malveira, D. M., de Sousa, D. L., Ribeiro, P. C. M., and Leite, A. C. S. (2015) Contribuição metodológica para análise dos benefícios na segurança do transito devido a modificações viárias. XIII Rio de Transportes, Rio de Janeiro, Brazil.
- Martin, J. L. (2002) Relationship between crash rate and hourly traffic flow on interurban motorways. *Accident Analysis and Prevention*, 34, 619-629.
- Mensah, A., and Hauer, E. (1998) ‘Two problems of averaging arising from the estimation of the relationship between accidents and traffic flow. *Transportation Research Record*, 1635, 37–43.
- Milton, J. C., Shankar, V. N., and Mannering, F. L. (2008) Highway crash severities and the mixed logit model: An exploratory empirical analysis. *Crash Analysis and Prevention*, 40, 260-266.
- Mitra, S., Chin, H. C., and Quddus, M. (2007) Study of Intersection Accidents by Maneuver

Type. Transportation Research Record, 1784, 43-50.

Montgomery, D. C., and Runger, G. C. (2014) Applied Statistics and Probability for Engineers. Wiley, USA.

Morris, M. D. (1991) Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics, 33(2), 161-174.

National Motorists Association (2018) State Speed Limit Chart.

Retrieved on May 2018 from <https://www.motorists.org/issues/speed-limits/>.

Nunes, D. F. (2012) Procedimento para Análise de Sensibilidade do Programa HDM-4. Master thesis, São Carlos School of Engineering, University of São Paulo, Brazil.

O'Donnell, C. J., and Connor, D. H. (1996) Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis and Prevention, 28 (6), 739-753.

Ozbay, K., Yang, H., Bartın, B., and Mudigonda, S. (2008) Derivation and validation of new simulation-based surrogate safety measure. Transportation Research Record, 2083, 105–113.

Pande, A., Abdel-Aty, M., and Hsia, L. (2005) Spatiotemporal variation of risk preceding crashes on freeways. Transportation Research Record, 1908, 26-36.

Penmetsa, P., and Pulugurtha, S. S. (2018) Modeling crash injury severity by road feature to improve safety. Traffic Injury Prevention, 19 (1), 102-109.

Piva, F. (2015) Fatores de Equivalência de Veículos Pesados em Rodovias de Pista Dupla. Master thesis, São Carlos School of Engineering, University of São Paulo, Brazil.

Poch, M., and Mannering, F. (1996) Negative Binomial Analysis of Intersection-Accident Frequencies. Journal of Transportation Engineering, 122, 05-113.

Potts, I. B., Harwood, D. W., Fees, C. A., and Bauer, K. M. (2014) Further development of the safety and congestion relationship for urban freeways SHRP 2 Report S2-L07-RR-3. Transportation Research Board, Washington, D.C.

Quddus, M. A., Wang, C., and Ison, S. G. (2010) Road traffic congestion and crash severity: Econometric analysis using ordered response models. Journal of Transportation Engineering, 136, (5), 424-435.

- Quddus, M. (2015) Effects of geodemographic profiles of drivers on injury severity from traffic crashes using multilevel mixed-effects ordered logit model. *Transportation Research Record*, 2514, 149–157.
- ROSPA (2017) Rural Road Safety. The Royal Society for the Prevention of Accidents. Retrieved on July 2017 from <https://www.rosipa.com/road-safety/advice/road-users/rural/>.
- Santner, T. J., Williams, B. J., and Notz, W. I. (2003) *The Design and Analysis of Computer Experiments*, Springer-Verlag, New York, NY.
- Saltelli, A., Chan, K., and Scott, M. (2000) *Sensitivity Analysis*. Wiley, New York, NY.
- Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004) *Sensitivity analysis in practice: A guide to assessing scientific models*. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, UK.
- Shahdah, U., Saccomanno, F., and Persaud, B. (2015) Application of traffic microsimulation for evaluating safety performance of urban signalized intersections. *Transportation Research Part C*, 60, 96-104.
- Spigolon, L. M. G. (2010) *Semáforo: Grupo focal convencional x grupo focal com informação do tempo de verde/vermelho restante*. Master thesis, São Carlos School of Engineering, University of São Paulo, Brazil.
- Stevanovic, A., Stevanovic, J., So, J., and Ostojic, M. (2015) Multi-criteria optimization of traffic signals: Mobility, safety and environment. *Transportation Research Part C*, 55, 46-68.
- TRB. (2010) *Highway Capacity Manual 2010*. Transportation Research Board of the National Academies, Washington, D.C.
- TRB (2016) *Highway Capacity Manual, Sixth Edition: A Guide for Multimodal Mobility Analysis*. Transportation Research Board, Washington, D.C.
- Trucano, T. G., Swiler, L. P., Igusa, T., Oberkampf, W. L., and Pilch, M. (2006) Calibration, validation and sensitivity analysis: What's what. *Reliability Engineering & System Safety*, 91, 1331-1357.
- Turner, S., Singh, R., and Nates, G. (2012) *Crash prediction models for signalized intersections: signal phasing and geometry*. NZ Transport Agency research report 483.
- Wang, X. S., and Abdel-Aty, M. (2006) Temporal and spatial analyses of rear-end crashes at

signalized intersections. *Accident Analysis and Prevention*, 38, 1137-1150.

Wu, Q., Chen, F., Zhang, G., Liu, X. C., Wang, H., and Bogus, S. M. (2014) Mixed logit model-based driver injury severity investigations in single and multi-vehicle crashes on rural two-lane highways. *Accident Analysis and Prevention*, 72, 105-115.

Xi, Y., Zhang, Y., and Liang, F. (2009) Crash Injury Severity Analysis Using Bayesian Ordered Probit Models. *Journal of Transportation Engineering*, 135 (1), 18-35.

Xie, Y., Zhao, K., and Huynh, N. (2012) Analysis of driver injury severity in rural single-vehicle crashes. *Accident Analysis and Prevention*, 47, 36-44.

Yamamoto, T., and Shankar, V. N. (2004) Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects. *Accident Analysis and Prevention*, 36 (5), 869-876.

Yang, W., L. Zhang, L. Zhang, and H. Shi. (2013) A Golden Ratio-Based Genetic Algorithm and Its Application in Traffic Signal Timing Optimization for Urban Signalized Intersections.

Yasmin, S., and Eluru, N. (2013) Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Crash Analysis and Prevention*, 59, 506-521.

Ye, F., and Lord, D. (2014) Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. *Analytic Methods in Accident Research*, 1, 72-85.

Younshik, C., Song, T. J., and Yoon, B. J. (2014) Injury severity in delivery-motorcycle to vehicle crashes in the Seoul metropolitan area. *Accident Analysis and Prevention*, 62 (5), 79-86.

Zhou, M., and Sisiopiku, V. P. (2014) Relationship Between Volume-to-Capacity Ratios and Accident Rates. *Transportation Research Record*, 1581, 47-52.