



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE TECNOLOGIA**  
**DEPARTAMENTO DE ENGENHARIA DE TRANSPORTES**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TRANSPORTES**

**LUCAS TITO PEREIRA SOBREIRA**

**REGIMES DE FLUXO VEICULAR E ACIDENTES DE TRÂNSITO EM**  
**INTERSEÇÕES SEMAFORIZADAS DE FORTALEZA**

**FORTALEZA**

**2019**

LUCAS TITO PEREIRA SOBREIRA

REGIMES DE FLUXO VEICULAR E ACIDENTES DE TRÂNSITO EM INTERSEÇÕES  
SEMAFORIZADAS DE FORTALEZA

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Transportes na Universidade Federal do Ceará, como requisito parcial para a obtenção do título de Mestre em Engenharia de Transportes. Área de Concentração: Planejamento e Operação de Sistemas de Transportes.

Orientador: Prof. Ph.D. Flávio José Craveiro Cunto

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- S659r Sobreira, Lucas Tito Pereira.  
Regimes de fluxo veicular e acidentes de trânsito em interseções semaforizadas de Fortaleza / Lucas Tito Pereira Sobreira. – 2019.  
132 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Transportes, Fortaleza, 2019.  
Orientação: Prof. Dr. Flávio José Craveiro Cunto.
1. Segurança viária. 2. Regimes de fluxo veicular. 3. Análise de agrupamentos. 4. Perturbações na corrente de tráfego. I. Título.

CDD 388

---

LUCAS TITO PEREIRA SOBREIRA

REGIMES DE FLUXO VEICULAR E ACIDENTES DE TRÂNSITO EM INTERSEÇÕES  
SEMAFORIZADAS DE FORTALEZA

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Transportes da Universidade Federal do Ceará, como requisito parcial para a obtenção do título de Mestre em Engenharia de Transportes. Área de Concentração: Planejamento e Operação de Sistemas de Transportes

Aprovada em 28/06/2019.

BANCA EXAMINADORA

---

Prof. Ph.D. Flávio José Craveiro Cunto (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Ph.D. Francisco Moraes de Oliveira Neto  
Universidade Federal do Ceará (UFC)

---

Prof. Ph.D. Felipe Caleffi  
Universidade Federal de Santa Maria (UFSM)

## AGRADECIMENTOS

À minha família, Clarice, Davi, Deusito, Dudu, Glauco, Maiara e Rafinha, por todo o carinho e constante incentivo durante minha caminhada acadêmica.

Ao professor e amigo Flávio Cunto, que foi meu orientador de Iniciação Científica, de Projeto de Graduação e de Mestrado; pela orientação, pelo incentivo durante toda minha vida acadêmica e por conselhos durante minha caminhada.

Aos professores do Departamento de Engenharia de Transportes da UFC que contribuíram para meu crescimento como estudante, em especial aos professores Felipe, Manoel, Moraes e Bruno.

Aos companheiros de trabalho, Marcelo Luna e Glauber, pela compreensão e pelas contínuas discussões e contribuições a esta Dissertação.

Aos membros do Grupo de Pesquisa em Segurança Viária, em especial a Sasaki, Caio, Vanessa e Gabriela, pelas contribuições e pelos pós-reunião.

Aos amigos de Mestrado, essencialmente aos membros da Obredet, pelas divertidas e longas noites de estudo compartilhadas.

À *Johns Hopkins University* e à *Bloomberg Philanthropies* por incentivarem o desenvolvimento de pesquisas em segurança viária e pela bolsa concedida ao longo do curso.

Aos avaliadores do presente trabalho, professores Francisco Moraes e Felipe Caleffi, pela atenção e prontidão que aceitaram participar da minha banca.

## RESUMO

O advento de novas tecnologias de controle e monitoramento de tráfego permite o desenvolvimento de estudos mais robustos na área da segurança viária, a partir da obtenção de dados de tráfego com curta agregação temporal (1 a 15 minutos), possibilitando a avaliação do efeito de variáveis com oscilações dentro de um mesmo dia (volume, velocidade, grau de saturação, ciclo semafórico) na frequência e severidade de acidentes de trânsito. Tradicionalmente, os estudos na área são baseados em entidades, com a avaliação da relação entre a frequência anual dos acidentes e as características médias anuais de interseções ou segmentos viários. Este trabalho, no entanto, a partir de dados desagregados, aplica uma abordagem embasada na condição de tráfego, agrupando as variáveis de tráfego por similaridade, em vez de aspectos médios de cada entidade. O uso de dados desagregados demanda um maior conhecimento sobre o horário reportado dos acidentes, visando à associação dos acidentes às condições precursoras de regimes de fluxo veicular (RFV). Um dos objetivos do trabalho consistiu no desenvolvimento de um algoritmo de detecção de perturbações na velocidade, a partir de dados de fiscalização eletrônica, com o intuito de comparar a diferença entre os horários das perturbações e dos horários que os acidentes foram reportados. Os resultados mostraram uma diferença média de 19 minutos entre os horários, além de apontar que acidentes com vítimas apresentaram menores diferenças médias em comparação aos acidentes com apenas danos materiais. Utilizando informações agregadas em intervalos de 15 minutos fornecidas pelo sistema *SCOOT*, consolidou-se um banco de dados com 4,1M RFV; os quais passaram por um processo de agrupamento de intervalos homogêneos. Os principais resultados da análise exploratória e da modelagem da relação entre os agrupamentos de RFV e os acidentes em interseções semaforizadas da zona urbana de Fortaleza foram: (i) a ocorrência de acidentes com apenas danos materiais é mais provável em regimes com altas demandas veiculares; (ii) a frequência de acidentes com vítimas em regimes de alta e de baixa demanda veicular é similar, indicando que uma parcela dos acidentes aconteceu em função da possibilidade de desenvolvimento de velocidade e outra devido à exposição e à vulnerabilidade dos usuários; (iii) os regimes que apresentaram as combinações de *baixa demanda veicular + ciclos semafóricos longos* e de *alta demanda veicular + ciclos semafóricos curtos* foram os mais críticos, principalmente em termos de acidentes com vítimas.

**Palavras-chave:** Segurança viária. Regimes de fluxo veicular. Análise de agrupamentos. Perturbações na corrente de tráfego.

## ABSTRACT

The advent of new technologies for monitoring and controlling traffic allows the development of more robust road safety studies, from traffic data collection at short temporal intervals (1 to 15 minutes), making possible the evaluation of the effect of variables that oscillate within the same day (vehicle volume, speed, degree of saturation, traffic light cycle) in the frequency and severity of traffic crashes. Traditionally, studies in road safety field are based on entities, assessing the relationship between the annual frequency of crashes and the average annual characteristics of intersections or road segments. However, this work was developed using disaggregated data, applying an approach based on the traffic condition, by grouping similar traffic variables, rather than annual average aspects of each entity. The use of disaggregated data requires a more reliable accuracy of crashes reported time, aiming a trustable association between the crash and its traffic precursor condition. One of the goals of this study was to develop a speed disturbance detection algorithm based on speed camera data, in order to compare the difference between the disturbance time and the time that crashes were reported. The results showed a 19 minutes average time difference and pointed out that crashes with victims presented smaller average differences in comparison to the crashes with only property damages. Using information aggregated in intervals of 15 minutes provided by the *SCOOT* system, a database with 4.1M traffic conditions was consolidated. A cluster analysis was performed to group similar traffic conditions. The main results of evaluation of the relationship between the clusters and the crashes at urban signalized intersections in Fortaleza were: (i) the occurrence of property damage only crashes is more likely in conditions with high vehicular demands; (ii) the frequency of crashes with victims in high and low vehicle demand conditions is similar, indicating that a portion of the crashes happened due to the possibility of speed development and another due to the exposure and the vulnerability of the users; (iii) the conditions that presented combinations of *low vehicle demand + long traffic cycles* and *high vehicular demand + short traffic cycles* were the most critical, especially in terms of crashes with victims occurrence.

**Keywords:** Road safety. Condition based approach. Cluster analysis. Traffic flow disturbances.

## LISTA DE FIGURAS

Figura 1 – Resumo das etapas da Dissertação .....	18
Figura 2 – Exemplo estudo entidade-anual e regime de fluxo veicular .....	20
Figura 3 – Resumo dos resultados obtidos por Imprialou et al. (2016). 1A: Dados agregados anualmente; 1B: Dados agregados em 15 minutos.....	25
Figura 4 – Exemplo de CURE plot.....	36
Figura 5 – Método de estimação do horário da ocorrência do acidente baseado em dados de velocidade de um laço à montante do acontecimento .....	37
Figura 6 – Método proposto .....	45
Figura 7 – Método proposto para análise entre tempo reportado da ocorrência de acidentes e perturbações na condição na velocidade em interseções semaforizadas .....	46
Figura 8 – Exemplo de aplicação do método para detecção de perturbações na velocidade .	48
Figura 9 – Método proposto para consolidação dos bancos de dados de RFV e dos acidentes .....	50
Figura 10 – Ilustração do posicionamento dos laços detectores do <i>SCOOT</i> .....	50
Figura 11 – Método proposto para definição dos agrupamentos de regimes de fluxo veicular	54
Figura 12 – Método proposto para modelagem por agrupamento e por entidade e para comparação dos resultados obtidos por tipo de abordagem .....	56
Figura 13 – Posição dos laços do <i>SCOOT</i> na cidade de Fortaleza .....	58
Figura 14 – Distribuição espacial dos equipamentos de fiscalização eletrônica de interseções semaforizadas e método de coleta dos acidentes .....	60
Figura 15 – Comparação de detecção com durações de três (A) e de um (B) minutos.....	61
Figura 16 – Exemplos de perturbações não detectadas pelo algoritmo: (A) colisão traseira; (B) colisão transversal.....	62
Figura 17 – Exemplos de perturbações detectadas pelo algoritmo, mas que não foram possíveis precisar os horários iniciais das oscilações: (A) colisão traseira; (B) colisão lateral .....	63
Figura 18 – Resumo das detecções das perturbações por etapa e por cenário .....	64
Figura 19 – Perturbações detectadas por dois laços em horários similares: (A) e (B) colisão transversal; (C) e (D) colisão traseira .....	66
Figura 20 – Perturbações detectadas por apenas um laço: (A) colisão transversal; (B) colisão lateral .....	67



Figura 21 – Exemplo de acidente reportado antes da detecção da perturbação na velocidade (colisão transversal) .....	68
Figura 22 – Interseções semaforizadas contempladas no estudo .....	70
Figura 23 – Boxplot e frequências das principais variáveis de tráfego por hora .....	73
Figura 24 – Distribuição horária dos acidentes totais e com vítimas .....	75
Figura 25 – Boxplot RFV associados aos acidentes por severidade .....	77
Figura 26 – Boxplot RFV associados aos acidentes por tipo .....	79
Figura 27 – Gráfico de dispersão e índice de correlação entre os atributos variáveis dos RFV .....	83
Figura 28 – Exemplo de configuração de agrupamentos de regimes de fluxo veicular .....	84
Figura 29 – Distribuição das amostras completa e reduzida .....	84
Figura 30 – Definição do número ótimo de agrupamentos: <i>Silhouette test</i> .....	85
Figura 31 – Definição do número ótimo de agrupamentos: <i>Gap Statistic test</i> .....	86
Figura 32 – Definição do número ótimo de agrupamentos: soma dos erros quadrados entre agrupamentos .....	87
Figura 33 – Configurações de agrupamentos selecionadas: <i>fluxo por faixa e velocidade</i> .....	88
Figura 34 – Frequência de acidentes por severidade a cada 10.000 horas por ARFV: fluxo por faixa e velocidade .....	91
Figura 35 – Matriz de correlação entre as variáveis dos RFV.....	93
Figura 36 – Frequência de acidentes por severidade a cada 10.000 horas por ARFV: <i>fluxo por faixa, velocidade e número de faixas</i> .....	95
Figura 37 – Frequência de acidentes por severidade a cada 10.000 horas por ARFV: <i>fluxo por faixa, velocidade e ciclos por hora</i> .....	96
Figura 38 – Análise complementar I: impacto do retrocesso temporal nas características de tráfego .....	98
Figura 39 – ARFV – acidentes com danos materiais: resíduos do modelo 1 .....	103
Figura 40 – ARFV – acidentes com danos materiais: representação gráfica do modelo 1 ....	104
Figura 41 – ARFV – acidentes com vítimas: resíduos do modelo selecionado .....	107
Figura 42 – ARFV – acidentes com vítimas: representação gráfica do modelo 10 .....	108
Figura 43 – Entidade-anual: matriz de correlação .....	109
Figura 44 – Entidades – acidentes com danos materiais: <i>CuRe plot</i> modelo 3 .....	111
Figura 45 – Entidades – acidentes com vítimas: <i>CuRe plot</i> modelo 6 .....	113
Figura 46 – Acidentes com danos materiais: comparação das estimativas por abordagem. (A) ARFV e (B) entidades-anual .....	114

Figura 47 – Entidades – acidentes com vítimas: representação gráfica do modelo 6 .....	115
Figura 48 – Acidentes com vítimas: comparação das estimativas por abordagem. (A) ARFV e (B) entidades .....	116

## LISTA DE TABELAS

Tabela 1 – Resumo de estudos de segurança viária em interseções semaforizadas .....	22
Tabela 2 – Resumo do estudo de caso para identificação dos critérios para aplicação do algoritmo de detecção de perturbações na velocidade .....	61
Tabela 3 – Estatísticas entre as diferenças entre o horário reportado dos acidentes e a perturbação na velocidade por cenário .....	67
Tabela 4 – Estatística descritiva dos regimes de fluxo veicular .....	71
Tabela 5 – Resumo do processo de obtenção dos acidentes e da associação aos RFV .....	74
Tabela 6 – Resumo dos acidentes associados a RFV por tipo e severidade.....	74
Tabela 7 – Estatística descritiva dos RFV associados a acidentes .....	76
Tabela 8 – Estatística descritiva das entidades-anuais .....	81
Tabela 9 – Resumo dos ARFV com $K = 8$ .....	90
Tabela 10 – Categorização das variáveis para definição das configurações de ARFV.....	93
Tabela 11 – ARFV – acidentes com danos materiais: Variáveis significativas por configuração aplicando modelagem de Poisson e Binomial Negativa .....	101
Tabela 12 – ARFV – acidentes com danos materiais: coeficientes dos modelos de distribuição Binomial Negativa.....	102
Tabela 13 – ARFV – acidentes com vítimas: Variáveis significativas por configuração aplicando modelagem de Poisson .....	105
Tabela 14 – ARFV – acidentes com vítimas: coeficientes do modelo de distribuição de Poisson .....	106
Tabela 15 – Entidades – acidentes com danos materiais: coeficientes e indicadores dos modelos com distribuição Binomial Negativa .....	110
Tabela 16 – Entidades – acidentes com vítimas: coeficientes e indicadores dos modelos com distribuição Binomial Negativa.....	112

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>14</b>
<b>1.2</b>	<b>Questões motivadoras .....</b>	<b>16</b>
<b>1.3</b>	<b>Objetivos.....</b>	<b>17</b>
1.3.1	<i>Objetivo geral.....</i>	17
1.3.2	<i>Objetivos específicos .....</i>	17
<b>1.4</b>	<b>Estrutura do trabalho .....</b>	<b>17</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA .....</b>	<b>19</b>
<b>2.1</b>	<b>Estudos de segurança viária desenvolvidos para interseções semaforizadas.....</b>	<b>20</b>
<b>2.2</b>	<b>Estudos que avaliam a relação entre regimes de fluxo veicular e acidentes .....</b>	<b>23</b>
2.2.1	<i>Estudos de regime de fluxo veicular macroscópico e acidentes de trânsito .....</i>	23
2.2.2	<i>Estudos de precursores de acidentes.....</i>	27
<b>2.3</b>	<b>Métodos para modelagem da frequência de acidentes.....</b>	<b>29</b>
2.3.1	<i>Métodos para modelagem: abordagem entidade-anual.....</i>	29
2.3.1.1	<i>Histórico das técnicas de calibração para modelagem da frequência de acidentes .....</i>	30
2.3.1.2	<i>Avaliação do ajuste do modelo e comparação entre modelos .....</i>	34
2.3.2	<i>Métodos para modelagem: abordagem regime de fluxo veicular macroscópico .....</i>	36
2.3.2.1	<i>Estimação do horário mais provável de ocorrência do acidente em função de perturbações na condição do fluxo veicular .....</i>	36
2.3.2.2	<i>Técnicas de agrupamentos de dados para definição de agrupamentos de regime de fluxo veicular .....</i>	38
2.3.2.3	<i>Técnicas de calibração e validação da modelagem de regime de fluxo veicular macroscópico .....</i>	42
<b>2.4</b>	<b>Considerações finais .....</b>	<b>43</b>
2.4.1	<i>Hipóteses a serem verificadas no trabalho .....</i>	43
<b>3</b>	<b>MÉTODO.....</b>	<b>45</b>
<b>3.1</b>	<b>Estimação do horário mais provável da ocorrência do acidente a partir de perturbações na velocidade da corrente de tráfego.....</b>	<b>46</b>
<b>3.2</b>	<b>Consolidação do banco de dados de regimes de fluxo veicular e de acidentes.....</b>	<b>49</b>
<b>3.3</b>	<b>Definição dos agrupamentos de regimes de fluxo veicular .....</b>	<b>53</b>

<b>3.4</b>	<b>Modelagem da relação entre agrupamentos de regimes de fluxo veicular e os acidentes e comparação entre os resultados obtidos com a modelagem por entidades e por agrupamentos .....</b>	<b>55</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>58</b>
<b>4.1</b>	<b>Estimação do horário mais provável da ocorrência do acidente a partir de perturbações na velocidade da corrente de tráfego.....</b>	<b>58</b>
4.1.1	<i>União do banco de dados de fiscalização eletrônica e acidentes .....</i>	59
4.1.2	<i>Estudo de caso para definição do critério de identificação de perturbação na velocidade .....</i>	59
4.1.3	<i>Análise das perturbações na velocidade e comparação com os horários reportados dos acidentes.....</i>	64
<b>4.2</b>	<b>Consolidação dos bancos de dados.....</b>	<b>69</b>
4.2.1	<i>Consolidação do banco de dados de regimes de fluxo veicular.....</i>	69
4.2.2	<i>Consolidação do banco de dados de acidentes associados aos regimes de fluxo veicular .....</i>	72
4.2.3	<i>Consolidação do banco de dados de acidentes associados às entidades-anuais .....</i>	80
<b>4.3</b>	<b>Definição dos agrupamentos de regimes de fluxo veicular .....</b>	<b>80</b>
4.3.1	<i>Amostragem segmentada dos regimes de fluxo veicular.....</i>	82
4.3.2	<i>Definição do número de agrupamentos baseados no fluxo por faixa e na velocidade .....</i>	85
4.3.3	<i>Definição das configurações dos agrupamentos de regimes de fluxo veicular .....</i>	92
4.3.4	<i>Análise complementar I: avaliação do impacto nas características dos RFV com a aplicação da correção média nos horários reportados dos acidentes.....</i>	98
<b>4.4</b>	<b>Modelagem estatística: relação entre os acidentes e as variáveis de tráfego, de operação e de geometria .....</b>	<b>99</b>
4.4.1	<i>Modelagem estatística: abordagem por agrupamentos de regimes de fluxo veicular .....</i>	100
4.4.1.1	<i>Modelagem por ARFV: acidentes com danos materiais .....</i>	100
4.4.1.2	<i>Modelagem por ARFV: acidentes com vítimas .....</i>	105
4.4.2	<i>Modelagem estatística: abordagem por entidades.....</i>	108
4.4.2.2	<i>Modelagem por entidades: acidentes com danos materiais.....</i>	110
4.4.2.2	<i>Modelagem por entidades: acidentes com vítimas .....</i>	111
4.4.3	<i>Comparação entre as modelagens estatísticas: ARFV e entidades .....</i>	113
<b>5</b>	<b>CONCLUSÃO.....</b>	<b>118</b>

<b>5.1</b>	<b>Limitações e trabalhos futuros .....</b>	<b>121</b>
	<b>REFERÊNCIAS .....</b>	<b>123</b>

## 1 INTRODUÇÃO

O principal objetivo dos estudos na área de segurança viária é desenvolver conhecimento sobre a relação da frequência e severidade dos acidentes de trânsito e de seus potenciais fatores contribuintes. Aproximadamente três décadas atrás, Hauer (1990) levantou a questão de que há a busca por vias mais seguras para os usuários, entretanto o conhecimento empírico dos tomadores de decisões relativo à prática de segurança viária é – em certos casos – inexistente.

A temática dos acidentes de trânsito vem sendo – ainda mais – discutida e contemplada no planejamento das cidades nos últimos anos, visto a necessidade de uma melhor compreensão desse fenômeno que é tratado como um problema de saúde pública. De acordo com a Organização Mundial da Saúde (OMS) (WHO, 2017), aproximadamente 1,25 milhão de pessoas morrem em acidentes por ano no mundo. Apesar de acumularem apenas metade da frota veicular mundial, os países de baixa e média renda concentram 90% dessas fatalidades. O relatório da OMS traz, ainda, que os acidentes viários são a principal causa das mortes de jovens com idade entre 15 e 29 anos no mundo.

Em âmbito nacional, o cenário também é preocupante. Em 2013, o Brasil foi o 57º país com maior taxa de mortalidade no trânsito por 100.000 habitantes no mundo (23,4), seis pontos acima da média mundial (WHO, 2015). Conforme dados do Seguro DPVAT, no Brasil ocorreram 41.151 mortes nas vias em 2017, consistindo em um aumento de 11% com relação ao ano anterior (LÍDER-DPVAT, 2017). Em esfera metropolitana, em 2016, a cidade de Fortaleza apresentou um índice de 10,7 mortes/100 mil habitantes, indicando uma redução em 11,7% com relação ao ano de 2015 e em 30% referente ao ano de 2011, época em que foi iniciada a campanha da Década da Ação para Segurança no Trânsito pela Organização das Nações Unidas (ONU) (AMC, 2017).

Contemplando os anseios de Hauer (1990), diversos estudos ressaltando a importância da utilização de indicadores do Desempenho da Segurança Viária (DSV) dentro do planejamento da mobilidade e da segurança viária das cidades foram desenvolvidos nas últimas décadas (HERMANS et al., 2009; WASHINGTON et al., 2006; WHO, 2015). Schulze e Koßmann (2010) dizem que para garantir uma mobilidade segura e reduzir a frequência de acidentes, um gerenciamento da segurança viária deve ser baseado em medidas mitigadoras fundamentadas por evidências científicas. Torres (2016) cita que a aplicação dos indicadores de DSV dentro do processo de tomada de decisão no planejamento de transportes pode reduzir significativamente os aspectos negativos em uma rede viária, impactando na

redução de gastos econômicos e sociais, reproduzido na contenção do número de acidentes e da necessidade de intervenções para melhorias na segurança viária.

Uma das ferramentas estatísticas mais utilizadas nos estudos de DSV são os Modelos de Previsão de Acidentes (MPA) que representam equações matemáticas calibradas e validadas com o intuito de relacionar a frequência de acidentes de trânsito com variáveis de exposição, de características de tráfego, de uso do solo, de atributos operacionais e geométricos das vias, entre outras. Uma importante parcela das pesquisas desenvolvidas utiliza variáveis de exposição e de tráfego com agregações temporais mensais ou anuais (CUNTO; CASTRO NETO; BARREIRA, 2011; HADAYEGHI; SHALABY; PERSAUD, 2003; HAUER, 2004; LYON; PERSAUD, 2002; SOBREIRA; CUNTO, 2014; WANG; QUDDUS; ISON, 2009). Entretanto é constatado que essa agregação temporal dos dados pode acarretar em modelos de menor utilidade, principalmente quando as variáveis explicativas possuem oscilação dentro de um mesmo dia, como a velocidade e o volume de veículos (DAVIS, 2004; IMPRIALOU et al., 2016).

Com o advento de novas tecnologias de controle e monitoramento do tráfego urbano, como laços indutores de detecção de veículos, fiscalização eletrônica e dados georreferenciados de GPS, viabilizou-se a obtenção de dados desagregados para a realização de modelagens mais robustas, visando uma melhor compreensão da natureza dos acidentes. Diversos estudos foram desenvolvidos objetivando relacionar a frequência e a severidade dos acidentes com variáveis de tráfego (velocidade média, volume veicular, densidade veicular, grau de saturação das vias, entre outras) utilizando agregações horárias e sub-horárias (GARBER; WU, 2001; IMPRIALOU et al., 2016; LORD; MANAR; VIZIOLI, 2005; STEMPEL et al., 2016; WANG et al., 2018). Apesar da maioria dos estudos evidenciarem uma relação linear ou exponencial positiva da velocidade média – principal variável investigada nas pesquisas – com os acidentes, não existe uma conclusão unânime sobre o assunto (HAUER, 2009). Existe um consenso, porém, de que o efeito da velocidade deve ser investigado considerando o impacto simultâneo com outras características do tráfego, por exemplo: volume e densidade veicular (AARTS; VAN SCHAGEN, 2006; GARBER; SUBRAMANYAN, 2001; LORD; MANAR; VIZIOLI, 2005).

A grande maioria das pesquisas realizadas com aplicação de variáveis de tráfego desagregadas e acidentes de trânsito foi realizada considerando situações de fluxo ininterrupto (rodovias). Para o ambiente urbano densamente semaforizado, como é o caso da cidade de Fortaleza/CE, que traz características de interrupção na corrente de tráfego sujeitando os



condutores a diversas paradas, aumentando a probabilidade da ocorrência de conflitos veiculares; verifica-se um déficit de trabalhos desenvolvidos sobre o tema.

Imprialou *et al.* (2016) sugerem um novo método para o desenvolvimento dos modelos a partir de agrupamentos de regimes de fluxo veicular (*condition-based*), em vez das entidades. Essa recente abordagem agrupa as informações (de tráfego, de características geométricas e operacionais das vias, de uso do solo, entre outras) da amostra de entidades em agrupamentos homogêneos de regimes de fluxo veicular; assim, cada acidente é associado a um grupo que contempla as condições anteriores à ocorrência do mesmo. Esse método será melhor abordado na seção 2.

Ao mesmo tempo em que a utilização de dados desagregados possibilita o desenvolvimento de análises mais robustas, surge a necessidade da identificação do horário mais provável da ocorrência do acidente, visto que os horários reportados pelos agentes responsáveis pelo atendimento do evento que constam nos bancos de dados podem ser imprecisos, prejudicando a identificação das condições de tráfego anteriores ao acidente (IMPRIALOU *et al.*, 2016; LEE; SACCOMANNO; HELLINGA, 2002; ZHENG, 2012).

## 1.2 Questões motivadoras

A partir da contextualização da problemática dos acidentes e dos estudos envolvendo regimes de fluxo veicular e segurança viária, torna-se possível o estabelecimento das seguintes questões motivadoras para desenvolvimento desta Dissertação:

- I) Como estimar o intervalo temporal mais provável da ocorrência de acidentes de trânsito em comparação aos horários reportados a partir da identificação de perturbações na condição do fluxo veicular?
- II) Como definir agrupamentos homogêneos de regimes de fluxo veicular que englobem características de tráfego e de geometria e operação das vias para estimação da relação das variáveis componentes dos agrupamentos com os acidentes de trânsito?
- III) Qual a relação entre os agrupamentos de regime de fluxo veicular e a frequência e severidade dos acidentes de trânsito em interseções semaforizadas de Fortaleza/CE?
- IV) Qual a diferença entre os resultados (variáveis significativas e coeficientes) obtidos com as modelagens por entidades e por agrupamentos de regime de fluxo veicular?

## **1.3 Objetivos**

### ***1.3.1 Objetivo geral***

O objetivo geral desta Dissertação é avaliar a relação entre os agrupamentos de regime de fluxo veicular com agregação sub-horária e a frequência e severidade dos acidentes de trânsito em interseções semaforizadas de Fortaleza/CE.

### ***1.3.2 Objetivos específicos***

Os seguintes objetivos específicos foram considerados no desenvolvimento desta Dissertação:

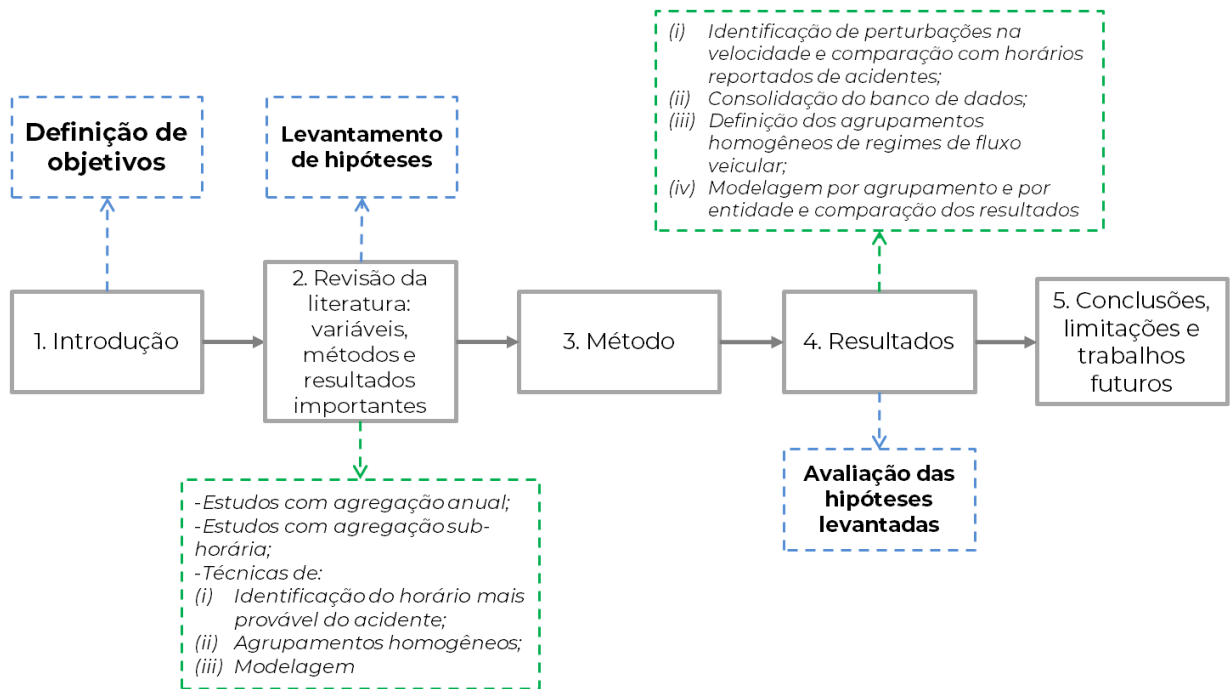
- I) Avaliar a relação entre o intervalo temporal da ocorrência de perturbações na condição de fluxo veicular e o horário reportado de acidentes de trânsito em Fortaleza/CE;
- II) Definir agrupamentos homogêneos de regimes de fluxo veicular que incluam características de tráfego e de geometria e operação das vias para estimação da relação das variáveis componentes do agrupamento com os acidentes de trânsito;
- III) Modelar a relação entre agrupamentos de regimes de fluxo veicular e a frequência e severidade dos acidentes de trânsito em interseções semaforizadas de Fortaleza/CE;
- IV) Efetivar uma análise comparativa entre os resultados obtidos com as modelagens por entidades e por agrupamentos de regime de fluxo veicular.

## **1.4 Estrutura do trabalho**

Este trabalho está estruturado em cinco capítulos. O capítulo presente apresentou a contextualização do tema a ser abordado, além das questões de motivadoras e dos objetivos geral e específicos. O capítulo 2 traz a revisão da literatura buscando expor as metodologias, os principais resultados e as limitações dos estudos desenvolvidos na temática levantada, apresentando, por fim, hipóteses levantadas sobre a relação entre as variáveis de tráfego e os acidentes de trânsito a serem avaliadas na seção dos resultados. O capítulo 3 exhibe o método proposto para alcançar os objetivos específicos. O capítulo 4 traz os resultados obtidos,

trazendo uma discussão com base nas hipóteses levantadas. Por fim, o capítulo 5 apresenta as conclusões, as limitações e as sugestões de trabalhos futuros. A Figura 1 traz um resumo das etapas deste trabalho.

Figura 1 – Resumo das etapas da Dissertação



Fonte: Elaborada pelo autor

## 2 REVISÃO BIBLIOGRÁFICA

A avaliação de indicadores de segurança viária tradicionalmente é realizada com o desenvolvimento de análises que consideram a frequência anual de acidentes de trânsito como medidas de desempenho associadas a características agregadas anualmente de tráfego e a aspectos de operação, de geometria e de uso do solo das entidades. Há também pesquisas que tratam os acidentes e as variáveis de tráfego em menores agregações temporais visando relacionar as características de regime do fluxo à ocorrência de acidentes em intervalos horários ou sub-horários. Os estudos desenvolvidos para interseções semaforizadas contemplam, em sua grande maioria, apenas o primeiro grupo de pesquisas citado acima, dificultando a avaliação da influência de variáveis que possuem oscilações durante o dia na frequência de acidentes.

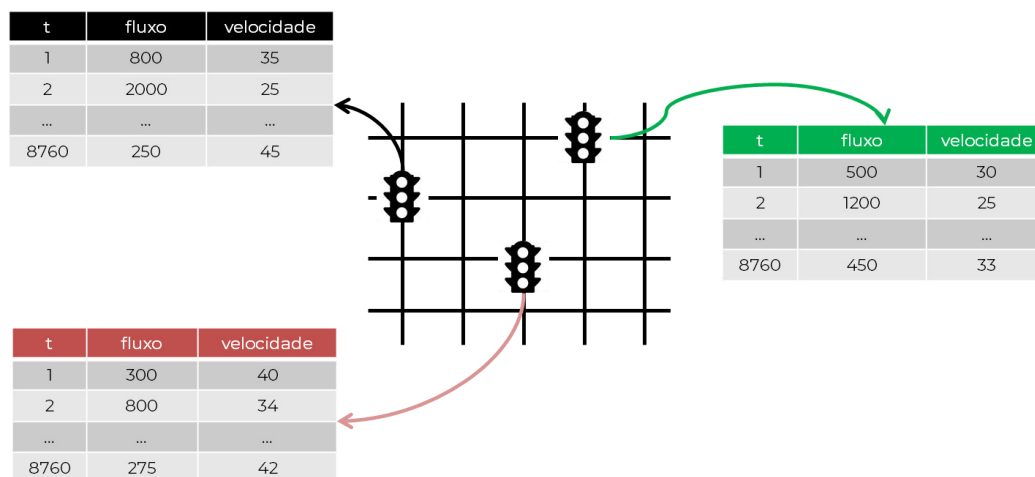
Para facilitar a condução dos termos ao longo do texto, convencionou-se que os estudos tradicionais com agregações anuais de acidentes e variáveis independentes são nomeados de entidade-anual e as pesquisas com utilização de dados agregados em hora ou intervalos de hora, de regimes de fluxo veicular (RFV). Apresenta-se, a seguir, um exemplo para consolidar a diferença entre as abordagens. Supondo uma amostra de três interseções semaforizadas em que se visa avaliar a relação entre o fluxo veicular e a velocidade, disponíveis em agregações de uma hora, e a frequência de acidentes em um determinado ano (Figura 2). Para o caso de dados desagregados, seriam estudados 26.280 RFV ( $365 \text{ dias} * 24 \text{ horas} * 3 \text{ interseções}$ ) cada um com suas informações de fluxo e velocidade, sendo associada a cada regime a ocorrência ou não de acidente naquela hora. Nessa abordagem não importa a qual interseção pertence o RFV, e sim, suas características. Destaca-se que os RFV também podem incluir atributos fixos de cada local, como geometria e operação da via e tipo de uso do solo.

A abordagem entidade-anual é baseada na associação de valores médios anuais de fluxo e velocidade para cada interseção ao somatório de acidentes que ocorreram ao longo de todo ano no local.

A revisão bibliográfica apresenta, inicialmente, os tipos de variáveis aplicadas e os principais resultados obtidos em estudos desenvolvidos para interseções semaforizadas predominantemente do tipo entidade-anual. Em seguida, estudos com foco nos RFV são mostrados, destacando as principais hipóteses levantadas para relação das variáveis de tráfego sub-horárias com os acidentes de trânsito para o ambiente de fluxo ininterrupto. Por fim, discute-se sobre os métodos adotados para modelagem em cada uma das abordagens,

saliendo as técnicas de calibração e validação, em ambos os casos, e os mecanismos para estimação do horário mais provável de ocorrência dos acidentes e para agrupamento das condições veiculares similares, na abordagem por RFV.

Figura 2 – Exemplo estudo entidade-anual e regime de fluxo veicular



Fonte: Elaborada pelo autor

## 2.1 Estudos de segurança viária desenvolvidos para interseções semaforizadas

A inclusão de indicadores de DSV no planejamento da mobilidade das cidades ao redor do mundo tem se tornado prática comum nas últimas décadas (WASHINGTON et al., 2006; WHO, 2015). O nível de agregação aplicado às possíveis variáveis explicativas relacionadas aos acidentes de trânsito está estritamente integrado ao tipo de planejamento que se deseja efetivar (GOMES et al., 2015). Técnicas macroscópicas, por exemplo, que consideram zonas de tráfego como entidades de estudo visam ao desenvolvimento de planejamentos estratégicos, buscando uma maior compreensão de características de uso do solo, da socioeconomia e de atributos operacionais e geométricos em relação à ocorrência de acidentes (GOMES et al., 2015; GOMES; CUNTO; SILVA, 2017; SIDDIQUI; ABDEL-ATY; CHOI, 2012; SOBREIRA; CUNTO, 2017; UKKUSURI et al., 2012). Análises microscópicas que, em geral, utilizam interseções semaforizadas, interseções não semaforizadas ou segmentos viários como entidades de estudo propõem-se a estimar relações de causalidade dos acidentes a partir de variáveis de exposição (fluxo de veículos e de pedestres) e de características da operação, geometria e uso do solo das entidades (GREIBE, 2003; HAUER, 2004; SOBREIRA; CUNTO, 2014; YANG; LOO, 2016).

A literatura apresenta inúmeros estudos desenvolvidos com foco em interseções semaforizadas (CASTRO; PALETI; BHAT, 2012; CHIN; QUDDUS, 2003; DONG et al., 2014; GUO; WANG; ABDEL-ATY, 2010; WONG; SZE; LI, 2007). Cunto *et al.* (2011), por exemplo, calibraram e validaram Modelos de Previsão de Acidentes (MPA) para 101 interseções semaforizadas da cidade de Fortaleza-BRA. Os autores consideraram variáveis de exposição (Volume Diário Médio Anual (VDMA) – via principal, VDMA – via secundária e VDMA – total) e de geometria (número de faixas, número de aproximações e presença de canteiro central). O modelo que apresentou melhores resultados foi composto pelas variáveis VDMA – total e número de faixas, verificando-se aumento na estimativa dos acidentes com o incremento de ambas as variáveis.

Xie *et al.* (2013) realizaram um estudo em 195 interseções semaforizadas ao longo de 22 corredores em áreas urbanas de Shangai-CHI. Variáveis como: número de aproximações, número de faixas, número de faixas de conversão, número de estágios semaforicos, duração do ciclo semaforico, grau de saturação, VDMA – total, velocidade média e variação da velocidade foram consideradas na análise. A velocidade foi medida através de dispositivos GPS em táxis durante uma semana no período de 13:00h-14:00h. Os principais resultados obtidos foram: (i) interseções com maior velocidade média estão relacionadas a maiores frequências de acidentes; (ii) entidades com quatro aproximações mostraram frequência de acidentes inferior às com três aproximações; (iii) o aumento no número de estágio semaforicos está atrelado ao aumento da quantidade de acidentes.

Abdel-Aty e Wang (2006) desenvolveram um estudo em 476 interseções semaforizadas localizadas na extensão de 41 corredores viários da Florida-EUA. As principais variáveis utilizadas na pesquisa foram: número de faixas (na via principal, secundária e somatório), número de faixas com conversão à direita, número de faixas com conversão à esquerda, quantidade de acessos em raios de 91 e 152 metros, distância da próxima interseção semaforizada, VDMA total, VDMA por faixa, limite de velocidade no corredor e tipo de uso do solo (residencial, comercial ou pouco adensado). Os resultados mostraram que interseções com elevados número de faixas, VDMA, limites de velocidade e próximas a outras interseções semaforizadas apresentaram relação com maiores frequências de acidentes. Já interseções com conversões à direita exclusivas, conversões à esquerda protegidas e localizadas em zonas pouco adensadas ou comerciais exibiram maiores níveis de segurança. A Tabela 1 traz o resumo de outros estudos focados em interseções semaforizadas.

Os estudos apresentados para interseções semaforizadas, em geral, agrupam anualmente a frequência dos acidentes associando-a a variáveis com agregações em níveis

mensais ou anuais. Esse tipo de abordagem permite a predição dos acidentes para determinadas condições, a identificação de pontos críticos e o maior conhecimento entre as relações de causa e efeito entre as variáveis. Todas as possíveis aplicações citadas anteriormente remetem a iniciativas reativas e que não buscam um maior detalhamento fenomenológico da ocorrência dos acidentes de trânsito.

Tabela 1 – Resumo de estudos de segurança viária em interseções semaforizadas

<b>Autores</b>	<b>Local</b>	<b>#Interseções</b>	<b>Variáveis significativas</b>
Chin e Quddus (2003)	Singapura	52	VDMA total, número de estágios por ciclo, número de conversões desprotegidas à esquerda e presença de fiscalização eletrônica
Wong <i>et al.</i> (2007)	Hong Kong, CHI	262	VDMA total, uso do solo, inclinação entre as aproximações e presença de passagem de trem
Guo <i>et al.</i> (2010)	Florida, EUA	170	VDMA via principal e secundária por faixa, número de faixas, número de conversões e presença de coordenação semaforica
Dong <i>et al.</i> (2014)	Tennessee, EUA	603	VDMA via principal e secundária, % de caminhões, número de conversões à esquerda e à direita e limite de velocidade

Fonte: Elaborada pelo autor

Nos últimos anos, com o desenvolvimento das tecnologias de controle e monitoramento de tráfego, ferramentas – principalmente para o ambiente rodoviário – vêm sendo elaboradas para proporcionar uma gestão dinâmica do fluxo veicular, visando à redução dos impactos negativos dos congestionamentos e à melhoria do desempenho da segurança viária, aspirando à identificação e, preferencialmente, à antecipação dos acidentes (USDOT, 2014). Por exemplo, o limite de velocidade variável é uma estratégia que busca a homogeneização da velocidade na corrente de tráfego, proporcionando maior fluidez e redução dos conflitos nas viagens (CALEFFI, 2018; YANG; LU; LIN, 2016).

No âmbito das vias de fluxo interrompido, as medidas de controle de tráfego estão, em sua maioria, concentradas na otimização em tempo real dos estágios semaforicos moderando os atrasos experimentados pelos condutores (CHILUKURI; PERRIN JR.; MARTIN, 2004; JHAVERI; PERRIN-JR; MARTIR, 2003). O surgimento de técnicas proativas em busca da melhoria da segurança viária em interseções semaforizadas depende de um maior conhecimento da interação entre as variáveis com oscilações intradiárias (volume de veículos, velocidade, número de paradas de veículos – gerando o movimento de parada-arranque –,

grau de saturação, entre outras) e os acidentes de trânsito (DAVIS, 2004; IMPRIALOU et al., 2016). Devido à escassa produção de estudos considerando interseções semaforizadas e aplicação de dados desagregados, na próxima seção são apresentadas pesquisas elaboradas a partir de RFV para rodovias, visando adquirir conhecimento das variáveis significativas, dos tipos de abordagens e dos resultados obtidos para esse tipo de regime, com o objetivo de facilitar o desenvolvimento das análises para o ambiente urbano altamente semaforizado.

## **2.2 Estudos que avaliam a relação entre regimes de fluxo veicular e acidentes**

O amplo desenvolvimento de estudos que visam relacionar variáveis de tráfego e a frequência e severidade dos acidentes constata a importância de um maior conhecimento sobre as condições antecessoras à ocorrência desses. Solomon (1964) apresentou o primeiro estudo que avalia essa relação a partir de dados de velocidade coletados em campo. O surgimento de novas tecnologias de monitoramento e controle de tráfego e de técnicas estatísticas possibilitou o avanço das pesquisas no tema que podem ser divididas em duas principais vertentes: (i) estudos de condições precursoras de acidente (*crash precursors*) que visam à detecção de perturbações na corrente de tráfego a fim de encontrar perfis relacionados à ocorrência de acidentes a partir de dados extremamente desagregados (um a cinco minutos) que antecederam ao acontecimento dos acidentes, e (ii) estudos de RFV macroscópicos que também buscam evidenciar a conexão entre padrões de tráfego e os acidentes, mas aplicando dados de maior agregação temporal (intervalos horários ou sub-horários).

### **2.2.1 Estudos de regime de fluxo veicular macroscópico e acidentes de trânsito**

Uma extensa gama de estudos considerando regimes de fluxo veicular e acidentes de trânsito foi desenvolvida nas últimas décadas. Esta seção apresenta as pesquisas segmentadas em três tipos de abordagem: (i) aplicação de variáveis médias que não representam o estado de fluxo veicular anterior ao acidente; (ii) associação dos acidentes a RFV antecessores ao seu acontecimento; (iii) avaliação das mudanças na segurança viária com alteração da velocidade média operacional e regulamentada nas vias.

O primeiro grupo de estudos - apesar de desagregadas em intervalos horários ou sub-horários - utilizou variáveis médias, as quais representam as condições típicas na via e não as anteriores à ocorrência do acidente (CEDER; LIVNEH, 1982; CHANG; XIANG, 2003; GARBER; WU, 2001; LORD; MANAR; VIZIOLI, 2005; MARTIN, 2002; SOLOMON,



1964; STEMPFEL et al., 2016; STUTER, 2004; WANG; QUDDUS; ISON, 2009; ZHOU; SISIOPIKU, 1997).

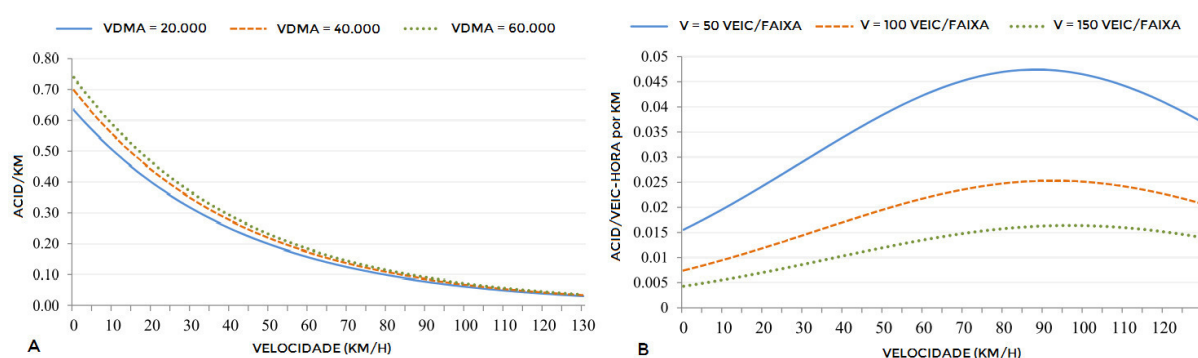
Lord *et al.* (2005), por exemplo, avaliaram a influência do volume, densidade e relação volume/capacidade (V/C) na frequência de acidentes em rodovias de Montreal-CAN. Os autores definiram três grupos de regimes: (i) segunda a sexta, (ii) sábado, (iii) domingo; caracterizando-os a partir de informações de laços detectores agregadas em intervalos horários obtidas em coletas realizadas de duas a três vezes ao ano, ou seja, representando condições médias e não precursoras ao acidente. Os resultados mostraram que o aumento da densidade, até certo ponto (15 a 20 veh/km/faixa), incrementa o risco de acidente, reduzindo em densidades mais elevadas. Esse resultado é corroborado por Zhou e Sisiopiku (1997), que também concluíram que a ocorrência de acidentes com vítimas diminui com o incremento no tráfego. Wang *et al.* (2009) mensuraram o efeito da variável índice de congestionamento, estimada em função da diferença entre o tempo de viagem observado e em fluxo livre, na frequência e severidade dos acidentes em 70 segmentos rodoviários na Inglaterra. Os autores verificaram uma tendência de redução dos acidentes com vítimas com o aumento do congestionamento, além de não observarem significância estatística da relação entre o índice e os acidentes totais.

Garber e Wu (2001) avaliaram a conexão entre variáveis de tráfego (velocidade, desvio padrão da velocidade, volume e ocupação) e 936 acidentes ocorridos em rodovias da Virginia-EUA. Os acidentes com horários reportados das 20h00m às 7h00m foram excluídos da análise, com o intuito de eliminar o possível efeito de iluminações precárias na segurança viária. Os pesquisadores agregaram informações fornecidas em intervalos de dez minutos em grupos de uma hora, estratificando os dias da semana em: (i) segunda; (ii) terça, quarta e quinta; (iii) sexta; (iv) sábado e domingo. O desvio padrão da velocidade foi estimado com base na variação entre as seis velocidades médias em cada hora, ponderadas pelo volume; portanto, altos desvios padrões indicam a transição do regime de tráfego (congestionado para fluxo livre, por exemplo) e não a variação individual da velocidade dos veículos dentro do intervalo de dez minutos. Os resultados evidenciaram que os acidentes estão positivamente relacionados com o volume e o desvio padrão da velocidade (i.e., períodos de transição de tráfego).

Por outro lado, alguns autores utilizaram bancos de dados mais complexos com disponibilidade de informações de regime de tráfego ao longo de todo o ano, possibilitando o relacionamento das reais características na movimentação veicular com os acidentes, em vez de medidas baseadas em tendências centrais (HAMZEHEI; CHUNG; MISKA, 2013;

IMPRIALOU et al., 2016; PEI; WONG; SZE, 2012; THEOFILATOS et al., 2018; ZHENG, 2012). Imprialou *et al.* (2016) desenvolveram um estudo sobre a relação entre a velocidade, o volume veicular e características geométricas e 9.310 acidentes em rodovias inglesas. Os pesquisadores realizaram uma comparação entre a modelagem considerando variáveis agregadas anualmente e em intervalos de 15 minutos, a Figura 3 traz o gráfico que resume os resultados encontrados com a análise feita para acidentes com vítimas (feridas e fatais).

Figura 3 – Resumo dos resultados obtidos por Imprialou et al. (2016). 1A: Dados agregados anualmente; 1B: Dados agregados em 15 minutos



Fonte: Adaptado de Imprialou *et al.* (2016)

Percebe-se que os produtos de cada modelagem foram significativamente diferentes. A análise com agregação anual (Figura 3A) mostra que a relação entre a velocidade e os acidentes com vítimas é inversamente proporcional e que o VDMA tem baixa relevância no estudo. Já o estudo desagregado apresenta conclusões mais intuitivas, em que o aumento da velocidade até certa faixa (80~90 km/h) está associado a maiores frequências de acidentes severos. Os autores levantaram a hipótese de que o desempenho de velocidades elevadas – acima de 100 km/h – está atrelado a vias de excelente infraestrutura, resultando na redução do risco de acidentes. A maior probabilidade de ocorrência de acidentes com vítimas em situações de baixo volume por faixa está vinculada – de acordo com os pesquisadores – a maiores variações na velocidade e ao incremento das interações veiculares (movimentos laterais). Os formatos das curvas obtidas na análise dos acidentes com danos materiais foram similares às geradas para acidentes com vítimas.

Na busca por uma metodologia de análise de eventos raros, Theofilatos *et al.* (2018) avaliaram a relação entre 17 acidentes de todas as severidades e variáveis de tráfego (velocidade, volume, densidade e porcentagem de caminhões) agregadas em uma hora obtidas a partir de três laços de detecção veicular em rodovia grega. Os autores utilizaram dados

completos de três anos e compararam condições com e sem a ocorrência de acidentes. Os resultados mostraram que baixas velocidades médias (50~60 km/h) estão atreladas a maiores potenciais de acidentes e que a porcentagem de caminhões não apresentou efeito significativo na segurança viária.

Zheng (2012) analisou a ocorrência de acidentes (N = 145) associada a três regimes de fluxo veicular determinadas com base na velocidade e na ocupação de rodovias: (i) tráfego congestionado; (ii) tráfego em transição; (iii) tráfego livre. Os resultados sugeriram probabilidades 1,62 e 6,61 maiores de ocorrência de acidentes em condições de transição e de congestionamento, respectivamente, quando comparadas a situações de fluxo livre. As conclusões obtidas por Zheng (2012) vão ao encontro das hipóteses avaliadas em outros estudos (GARBER; WU, 2001; STEMPFEL et al., 2016; TANISHITA; VAN WEE, 2017).

Na tentativa de transferir os resultados encontrados nos estudos supracitados para as interseções semaforizadas, supõe-se que RFV com maiores volumes veiculares podem estar associados a maiores frequências de acidentes, principalmente aqueles com apenas danos materiais, visto que a presença do semáforo, em conjunto com a alta exposição, incrementa as interações veiculares, sobretudo as longitudinais, podendo aumentar as probabilidades de ocorrência de acidentes. Outro fator que pode ser importante para entendimento do fenômeno é a duração do ciclo semafórico, que está relacionada a maiores ou menores quantidades de interações veiculares – tanto longitudinais quanto transversais – no cruzamento, podendo-se levantar a hipótese de que RFV com ciclos mais curtos, que podem estar vinculados a maiores frequências de conflitos veiculares, estão associados a probabilidades mais altas de ocorrência de acidentes. O ambiente urbano adensado de semáforos é caracterizado pela presença de movimentações não motorizadas, que são menos frequentes no ambiente rodoviário, podendo impactar no efeito da velocidade na frequência de acidentes com usuários vulneráveis.

Uma terceira frente de pesquisas busca mensurar o efeito de mudanças nas velocidades médias na segurança viária (ELVIK, 2009; ELVIK; CHRISTENSEN; AMUNDSEN, 2004; HAUER, 2009; NILSSON, 2004). Elvik *et al.* (2004) desenvolveram o *Power Model* – que consiste em uma formulação matemática visando relacionar mudanças na velocidade com alterações na frequência de acidentes por severidade – a partir da meta-análise de 98 diferentes estudos realizados em ambientes rurais e urbanos focados em pesquisas do tipo antes e depois. A função exponencial do *Power Model* indica que a redução na velocidade está altamente atrelada a melhorias na segurança viária, por exemplo: (i) a mudança da velocidade média de 70 para 63 mph está associada à redução em 48% dos acidentes fatais;

(ii) a diminuição da velocidade média de 30 para 27 mph está relacionada ao decréscimo médio de 44% das fatalidades.

Entretanto, ao avaliar a credibilidade do *Power Model*, Hauer (2009) destaca que resultados do modelo não são, em sua totalidade, evidenciados pelos dados analisados. O autor elenca alguns pontos para realçar a dificuldade na avaliação do impacto de alterações na velocidade na segurança viária: (i) não há a distinção entre veículos que realmente reduziram a velocidade daqueles que estão realizando conversões; (ii) o tipo de dado geralmente aplicado (sem segregação por severidade) não permite afirmar se houve redução nas taxas de ocorrência de acidentes ou na severidade dos sinistros (o que também afeta as taxas de reportabilidade, visto que acidentes mais severos tendem a ser mais reportados); (iii) a velocidade desempenhada na via e a velocidade da ocorrência do acidente não são necessariamente as mesmas. Entretanto, Hauer (2009) afirma que se a velocidade média aumenta enquanto outras condições (veículos, infraestrutura, serviços médicos, entre outras) permanecem as mesmas, os acidentes serão mais severos e, conseqüentemente, mais reportados.

Embora parte dos resultados conduza a hipóteses de que a probabilidade de ocorrência de acidentes aumenta em condições de congestionamento (baixa severidade) e de transição (maior severidade), não há evidências suficientes na literatura que garantam as assertivas. Existe, porém, consenso de que o efeito das variáveis de tráfego deve ser investigado simultaneamente, ou seja; velocidade, volume e densidade veicular devem ser analisados em conjunto e não separadamente (AARTS; VAN SCHAGEN, 2006; GARBER; SUBRAMANYAN, 2001; LORD; MANAR; VIZIOLI, 2005).

Ressalta-se que os estudos apresentados foram desenvolvidos para o ambiente de fluxo ininterrupto (rodovias), existindo clara lacuna na aplicação de pesquisas com dados desagregados de fluxo veicular para meios urbanizados densamente semaforizados, em que há a predominância do fluxo interrompido, sujeitando os condutores a diversos movimentos do tipo parada-arranque e – conseqüentemente – aumentando a probabilidade de conflitos veiculares.

### ***2.2.2 Estudos de precursores de acidentes***

Os estudos de precursores de acidentes são baseados na utilização de dados de fluxo veicular altamente desagregados (um a cinco minutos) obtidos em consecutivos laços detectores e têm como objetivo a identificação de perturbações ou condições atípicas na

corrente de tráfego que possam estar relacionadas à ocorrência de acidentes, vislumbrando – no melhor cenário – o desenvolvimento de medidas proativas em tempo real que reduzam a probabilidade de acontecimento do acidente (ABDEL-ATY et al., 2012; ABDEL-ATY; PANDE, 2005; ABDEL-ATY; PEMMANABOINA, 2006; HOURDOS et al., 2006; HUANG et al., 2016; LEE; ABDEL-ATY; HSIA, 2006; LEE; HELLINGA; SACCOMANNO, 2003; OH et al., 2005; SHI; ABDEL-ATY, 2015). Em geral, as variáveis avaliadas nesses estudos foram: velocidade absoluta, variação da velocidade no trecho e entre faixas, diferença da velocidade nos laços à montante e à jusante, densidade veicular e ocupação média.

O termo precursores de acidentes (*crash precursors*) foi idealizado por Lee *et al.* (2002) que desenvolveram uma análise com dados agregados em cinco minutos de 38 laços detectores em um trecho de 10 km rodoviários de Toronto-CAN. As principais conclusões dos autores foram: (i) o aumento da variação da velocidade no trecho e entre faixas está relacionado ao incremento de acidentes; (ii) o acréscimo na densidade veicular está atrelado a maiores riscos de acidentes; (iii) trechos com possibilidade de entrada e saída de veículos (aumento de movimentação lateral) estão associados a maiores frequências de acidentes do que trechos “diretos” (*straight sections*).

Pirdavani *et al.* (2015) realizaram um estudo em 42,5 km de segmentos rodoviários belgas a partir de informações obtidas em um minuto agregadas em intervalos de 5, 10 e 15 minutos antecessores ao acidente. A ausência de comprovação na literatura de qual a melhor agregação a ser utilizada motivou a investigação por parte dos autores, visto que Abdel-Aty *et al.* (2012) evidenciaram que a aplicação de dados agregados em um minuto está sujeita a “ruídos” ocasionados pelo acidente. Os pesquisadores belgas adotaram uma metodologia de comparar as variáveis de tráfego antecessoras a momentos de acidente e de não-acidente, considerando o intervalo de uma hora do horário reportado da ocorrência. A analogia foi baseada na relação de 1:4, em que um intervalo de acontecimento de acidente foi comparado a quatro de não-acidente. Os resultados mostraram que (i) baixas densidades veiculares estão associadas a menores riscos de acidentes; (ii) o aumento do desvio padrão da velocidade no laço à montante está atrelado a maiores probabilidades de incidentes; (iii) o incremento da diferença nas velocidades à montante e à jusante está relacionado a maiores frequências de acidentes. O modelo calibrado pelos autores – que apresentou estatísticas mais significativas com a agregação de cinco minutos – conseguiu um índice de previsão da ocorrência de acidentes de 70%, além de prever 90% das condições de não-acidente (falso positivo = 10%).

Abdel-Aty e Pande (2005), Kwak e Kho (2016) e Huang *et al.* (2017) também utilizaram o método de comparação entre binários acidente/não-acidente no desenvolvimento

de modelos de previsão de acidentes em tempo real, obtendo em média 70% de acerto na detecção dos acidentes.

Roshandel *et al.* (2015) realizaram um estudo de revisão bibliográfica baseado em 13 pesquisas que avaliaram precursores de acidentes em rodovias. Os resultados da meta-análise corroboram com os achados apresentados anteriormente em que os aumentos na variação da velocidade e no volume veicular estão associados a maiores probabilidades de acidentes, enquanto maiores velocidades e – consequentemente – menores volumes estão atrelados a condições com menores frequências de acidentes. É importante destacar que, mesmo com a utilização de dados altamente desagregados (um a cinco minutos), as conclusões não devem ser generalizadas ao nível de indivíduo, mantendo-se a referência a condições de tráfego (ZHENG, 2012).

Transportando os achados das rodovias para o ambiente de interseções semaforizadas, espera-se que fatores como velocidade média, desvio padrão da velocidade, *headway* médio e desvio padrão do *headway* estejam relacionados à ocorrência dos acidentes, visto que a característica de interrupção na corrente de tráfego ocasionada pela presença do semáforo faz com que a distância entre veículos e a velocidade desenvolvida pelos condutores sejam, possivelmente, fatores essenciais para entendimento do fenômeno aplicando curtas agregações temporais.

## **2.3 Métodos para modelagem da frequência de acidentes**

Esta seção traz os métodos aplicados em estudos de frequência de acidentes nas abordagens entidade-anual e RFV macroscópico. São apresentadas características dos bancos de dados, técnicas de calibração e validação dos modelos e peculiaridades para cada uma das abordagens.

### ***2.3.1 Métodos para modelagem: abordagem entidade-anual***

O banco de dados para a abordagem entidade-anual é constituído da associação da frequência anual (ou mensal) de acidentes às características médias de tráfego e aos aspectos de operação, de geometria e de uso do solo de cada entidade. Devido à extensa literatura produzida na área, as técnicas de calibração dos modelos passaram por considerável processo evolutivo nas últimas décadas. Esta seção é dividida em um breve histórico dessas técnicas

aplicadas em estudos entidade-anual com destaque em suas vantagens e limitações e na apresentação de técnicas de avaliação do ajuste e comparação entre modelos.

### 2.3.1.1 Histórico das técnicas de calibração para modelagem da frequência de acidentes

Os primeiros estudos desenvolvidos aplicaram modelos baseados na distribuição Normal com estimação dos coeficientes a partir do método dos mínimos quadrados. Por assumir variáveis dependentes contínuas e – principalmente – demandar homoscedasticidade (variância dos erros constantes para todo o conjunto de variáveis independentes); esse tipo de modelo não é adequado para estudo dos acidentes de trânsito, que representam um fenômeno raro, aleatório, discreto e não negativo (HAUER; HAKKERT, 1988; LORD; MANNERING, 2010). Buscando lidar com as características apresentadas dos acidentes, a elaboração de Modelos Lineares Generalizados (MLG) assumindo a distribuição de Poisson foi o ponto de partida para o desenvolvimento de análises estatisticamente mais confiáveis (JOSHUA; GARBER, 1990; JOVANIS; CHANG, 1986). A Equação 1 traz a formulação geral dos MLG.

$$y_i = \alpha * \left[ \prod_i A_i^{\beta_i} \right] * e^{\sum_j (\gamma_j B_j)} \quad (1)$$

Em que:  $y_i$  representa o número estimado de acidentes na entidade  $i$ ;  $A$  e  $B$  são vetores de variáveis independentes;  $\alpha$ ,  $\beta$  e  $\gamma$  são vetores de coeficientes do modelo.

A distribuição de Poisson é aplicada na obtenção da probabilidade de ocorrência em determinado intervalo de tempo de variáveis aleatórias, discretas e não negativas. A probabilidade de Poisson, diferentemente da Normal, tem variância igual à média e não exige homoscedasticidade entre as variáveis; possibilitando a incorporação de certo caráter disperso dos acidentes. A formulação da distribuição de Poisson é apresentada na Equação 2. Os parâmetros dos modelos de Poisson podem ser estimados com o método da maximização da verossimilhança que, em outras palavras, a partir do produtório das probabilidades da ocorrência de  $Y$  acidentes em determinada entidade  $i$  associadas a parâmetros  $\beta$  e variáveis explicativas  $X$  (Equação 3) maximiza a probabilidade dos dados se ajustarem à distribuição do modelo (Equação 5) utilizando a função de ligação logarítmica (Equação 4) para facilitar o desenvolvimento das formulações matemáticas.

$$P(y_{it}) = \frac{\lambda_{it}^{y_{it}} * \exp(-\lambda_{it})}{y_{it}!} \quad (2)$$

$$L(\beta) = \prod_{i=1}^n f_i(Y_i) = \prod \frac{[\mu(X_i, \beta)]^{Y_i} \exp[-\mu(X_i, \beta)]}{Y_i!} \quad (3)$$

$$\ell = \ln[L(\beta)] \quad (4)$$

$$\frac{\partial \ell}{\partial \beta} = 0 \quad (5)$$

Em que:  $P(y_{it})$  representa a probabilidade de ocorrência de  $y$  acidentes na entidade tipo  $i$  no período  $t$ ;  $\lambda_{it}$  é a taxa média de acidentes na entidade tipo  $i$  no período  $t$ ;  $Y_i$  é o vetor de acidentes observados em cada entidade;  $X_i$  é o vetor de variáveis independentes;  $\beta$  é o vetor com as estimativas dos parâmetros para cada variável independente.

Apesar de incorporar certa flutuação dos acidentes ao longo do tempo nas entidades, em alguns casos a dispersão acontece em níveis tão elevados (conhecida como super-dispersão) que a distribuição de Poisson não é capaz de lidar. Em outras palavras, a super-dispersão representa a existência de entidades que possuem atributos (variáveis independentes) similares, mas com diferentes taxas médias de acidentes. Assim, foi proposta a adaptação dos modelos de Poisson com inclusão de um parâmetro de superdispersão ( $\phi$ ) (HAUER, 2001).

A conhecida e amplamente utilizada distribuição Binomial Negativa (ou Poisson-gama) incorpora o parâmetro  $\phi$  assumindo distribuição gama com média igual a um e variância igual a  $1/\phi$  (CUNTO; CASTRO NETO; BARREIRA, 2011; HADAYEGHI; SHALABY; PERSAUD, 2007; KIM; WASHINGTON, 2006; LORD; BONNISON, 2007; PERSAUD, 1993; SOBREIRA; CUNTO, 2014). Com isso, as formulações tanto da função de probabilidade (Equação 6) quanto da estimação dos parâmetros (Equação 7) são alteradas em relação à Poisson. Destaca-se que o parâmetro da super-dispersão é estimado simultaneamente com os parâmetros  $\beta$  do modelo.

$$Y_{it} \sim BN(\lambda_{it}, \lambda_{it} + \frac{\lambda_{it}^2}{\phi}) \quad (6)$$

$$L(\beta) = \prod_{i=1}^n f_i(Y_{it} = y_{it}, \phi, \lambda_{it}) = \prod \frac{\Gamma(y_{it} + \phi)}{\Gamma(\phi) y_{it}!} \left( \frac{\phi}{\lambda_{it} + \phi} \right)^\phi \left( \frac{\lambda_{it}}{\lambda_{it} + \phi} \right)^{y_{it}} \quad (7)$$



Em alternativa aos modelos com distribuição Binomial Negativa (BN), recentemente modelos com distribuição Poisson-lognormal foram aplicados em estudos de segurança viária (AGUERO-VALVERDE; JOVANIS, 2008; LORD; MIRANDA-MORENO, 2008). A Poisson-lognormal apresenta estrutura similar à BN com diferenciação na distribuição do parâmetro da super-dispersão que segue a probabilidade log-normal, em vez da gama. Lord e Mannering (2010) sugerem que os modelos Poisson-lognormal são mais flexíveis que os BN, entretanto exibem formulação de maior complexidade.

Ao longo do histórico processo de desenvolvimento da modelagem de acidentes de trânsito, alguns pesquisadores observaram que em seus bancos de dados a considerável presença de entidades com frequência de acidentes igual a zero tornara a aplicação dos modelos de Poisson ou BN inviável, visto que essas não conseguiam estimar parâmetros adequados para lidar com os “excessos de zeros” na variável dependente (GARBER; WU, 2001; LAMBERT, 1992). Assim, adaptações aos modelos de Poisson e BN foram desenvolvidas para incorporação dos “excessos de zeros” na modelagem gerando as regressões Poisson Inflada de Zero (PIZ) (*Zero Inflated Poisson*) e Binomial Negativa Inflada de Zero (BNIZ) (*Zero Inflated Negative Binomial*).

Os modelos inflados de zero não são estruturados como modelos de contagem e sim segregando em probabilidades binárias de evento = 0 (não acidente) e evento > 0 (ocorrência de acidente). Dada a característica rara dos acidentes de trânsito, esses tipos de modelos foram amplamente utilizados em estudos de segurança viária (GARBER; WU, 2001; KUMARA; CHIN, 2003; LEE; MANNERING, 2002; QIN; IVAN; RAVISHANKER, 2004; SHANKAR et al., 2003).

Lord *et al.* (2005) desenvolveram um estudo misto empregando meta-análise e testes empíricos para avaliar o uso das distribuições de Poisson, PIZ e BNIZ na modelagem de acidentes. Nas pesquisas analisadas, foi verificado que a aplicação dos modelos inflados de zero ocorreu quando o número de entidades sem ocorrência de acidentes foi de 2 a 12% superior do que seria estimado a partir da distribuição de Poisson e da média de acidentes por ano. Os autores também destacam que a utilização desse tipo de modelo, fenomenologicamente, não é ideal; visto que a preponderância de zeros não é um indicativo de que os acidentes representam um evento binário. Quatro motivos são citados como possíveis motivadores do excesso de zeros nas análises: (i) combinação de entidades heterogêneas com altas variações na exposição e risco aos acidentes; (ii) amostras espacialmente ou temporalmente pequenas; (iii) baixa taxa de reportabilidade dos acidentes; (iv) omissão de importantes variáveis explicativas no modelo.

Uma das possibilidades para lidar com o “excesso de zeros” é a incorporação de séries históricas de acidentes para calibração dos modelos. Se por um lado a utilização de mais anos de ocorrência de acidentes aumenta o tamanho amostral, reduzindo as entidades com zero acidente e aumentando a média dos incidentes; por outro lado inclui-se um caráter longitudinal ao experimento que não pode ser negligenciado (LORD; PERSAUD, 2000; MOUNTAIN; MAHER; FAWAZ, 1998). O caráter longitudinal corresponde a possíveis efeitos na segurança viária ocasionados por evoluções tecnológicas dos veículos, desenvolvimento de políticas públicas, aumento da fiscalização no trânsito, mudança na reportabilidade ou estrutura de coleta dos acidentes, entre outros.

A modelagem tradicional não é capaz incorporar o efeito de tendências temporais aos modelos, surgindo como alternativa a aplicação de constantes temporais (MOUNTAIN; MAHER; FAWAZ, 1998; SOBREIRA; CUNTO, 2014) ou a utilização de Equações de Estimção Generalizada (EEG) (BALLINGER, 2004; LORD; PERSAUD, 2000; MOHAMMADI; SAMARANAYAKE; BHAM, 2014; SOBREIRA; CUNTO, 2014; WU et al., 2014). As EEG são uma extensão dos MLG nas quais é introduzida a variabilidade temporal dos acidentes dentro de uma mesma entidade. Para modelagem das EEG é necessária a definição da estrutura de correlação temporal da variável dependente que pode ser: (i) intercambiável; (ii) auto-regressiva; (iii) não estruturada; (iv) independente. De modo geral, os estudos avaliados passaram a incorporar o fator temporal quando séries históricas de pelo menos cinco anos foram utilizadas.

Para o desenvolvimento de análises que buscam segregar os acidentes por severidade, as técnicas de calibração apresentadas nos últimos parágrafos que lidam com apenas uma variável dependente (análise univariada) exibem certas limitações: (i) interdependência entre os acidentes por severidade pode existir (por exemplo: a frequência de fatalidades possivelmente não sofre reduções ou aumentos sem que os acidentes com feridos e com danos materiais sejam afetados); (ii) aspectos da geometria, da operação e do uso do solo das vias podem afetar a severidade dos acidentes em diferentes níveis (LORD; MANNERING, 2010; MA, 2006). Com isso, surge a modelagem multivariada que considera mais de uma variável dependente e estima os parâmetros a partir de técnicas de estatística bayesiana, possibilitando a modelagem dos acidentes segregados por severidade (AGUERO-VALVERDE; JOVANIS, 2009; EL-BASYOUNY; SAYED, 2009; JONATHAN; WU; DONNELL, 2016; MA, 2006; MA; KOCKELMAN; DAMIEN, 2008; PARK; LORD, 2008).

Outras técnicas, como: modelo de *Conway-Maxwell-Poisson*, modelo gama, modelo generalizado aditivo, modelo de efeitos aleatórios, modelo multinomial negativo, modelo de

parâmetros aleatórios, modelos hierárquicos e rede neurais são outras das abordagens verificadas em estudos da modelagem da frequência de acidentes de trânsito (LORD; MANNERING, 2010).

### 2.3.1.2 Avaliação do ajuste do modelo e comparação entre modelos

Devido à incerteza da influência de cada variável independente na representação do fenômeno dos acidentes, geralmente vários modelos são calibrados com aplicação de diversas combinações entre as variáveis. Assim, a definição de indicadores para a avaliação do modelo mais ajustado é necessária. Como a maioria dos métodos de estimação dos parâmetros dos modelos são baseados na maximização da verossimilhança, o valor do logaritmo da verossimilhança é um dos mais comuns indicadores utilizados para medir o ajuste dos dados à distribuição de probabilidade escolhida, em que maiores valores estão relacionados a melhores ajustes (ABDEL-ATY et al., 2005; CHEN; SHEN, 2016; CUNTO; CASTRO NETO; BARREIRA, 2011; HAQUE; CHIN, 2010; HOSSAIN; MUROMACHI, 2013; SOBREIRA; CUNTO, 2014; TORRES, 2016).

Outra métrica bastante utilizada para comparação entre modelos é o critério de informação de Akaike (*Akaike Information Criteria – AIC*) (AKAIKE, 1974). O *AIC* compara modelos baseando-se no equilíbrio entre o viés e a variância que a inclusão de variáveis traz aos mesmos. Por exemplo: modelos com menos variáveis tendem a possuir menor viés e maior variância, já a inclusão de mais variáveis está relacionada ao aumento do viés do modelo (incorporação de mais características específicas dos locais e dos anos estudados, diminuindo a transferibilidade espacial e temporal do modelo) e à redução da variância. O modelo que apresenta o menor valor de *AIC* é considerado o mais equilibrado (CUNTO; CASTRO NETO; BARREIRA, 2011). A Equação 8 traz a formulação do *AIC*.

$$AIC = -2 \log(L) + 2p \quad (8)$$

Em que: *AIC* representa o critério de informação de Akaike; *L* é o máximo valor da função de verossimilhança do modelo estimado; *p* expressa o número de parâmetros do modelo.

Outro método para avaliar o ajuste dos modelos é a validação cruzada, que consiste na verificação dos resíduos obtidos com a aplicação dos modelos em um banco de dados auxiliar (outro ano de acidentes ou percentual do banco de dados consolidado não utilizado na

calibração) ou em porção aleatória da amostra utilizada na calibração. Shi e Abdel-Aty (2015) utilizaram 30% dos 243 acidentes estudados como amostra de validação da modelagem de variáveis de tráfego em tempo real. Cunto, Sobreira e Ferreira (2013) empregaram 50% das 89 interseções semaforizadas para validação da transferibilidade dos parâmetros estimados dos MPA no *Highway Safety Manual* para a cidade de Fortaleza-BRA.

Além das medidas apresentadas, pode-se também medir os resíduos dos modelos em âmbito global com aplicação dos indicadores do Erro Médio Absoluto (EMA – Equação 9) e da Raiz do Erro Quadrático Médio (REQM – Equação 10) (CUNTO; SOBREIRA; FERREIRA, 2013; GÓMEZ; BOCAREJO, 2015; HADAYEGHI, 2009; JONSSON et al., 2009; SOBREIRA; CUNTO, 2017).

$$EMA = \frac{\sum_i^n |\hat{y}_i - y_i|}{n} \quad (9)$$

$$REQM = \sqrt{\frac{\sum_i^n (\hat{y}_i - y_i)^2}{n}} \quad (10)$$

Em que:  $\hat{y}_i$  representa o estimado pelo modelo;  $y_i$  é o valor observado;  $n$  é o número de observações.

Entretanto, a verificação exclusiva dos resíduos em âmbito global pode não ser o suficiente para comprovação do ajuste do modelo em todo o conjunto amostral (HAUER, 2015). Assim, para medição dos resíduos locais da modelagem, Hauer e Bamfo (1997) propuseram a utilização do gráfico de resíduos acumulados (*CUMulative RESiduals – Cure plot*) para avaliação do ajuste das estimações compreendendo todo o intervalo das variáveis independentes. O *CURE plot* é baseado na plotagem dos resíduos acumulados pelo modelo para cada observação e na definição de duas envoltórias composta pelo valor de  $\pm 2\sigma^*$  (Equação 11). É esperado para um modelo bem ajustado que os resíduos acumulados variem em torno do eixo das abscissas e estejam sempre dentro das duas envoltórias. A Figura 4 apresenta um exemplo de *CURE plot*.

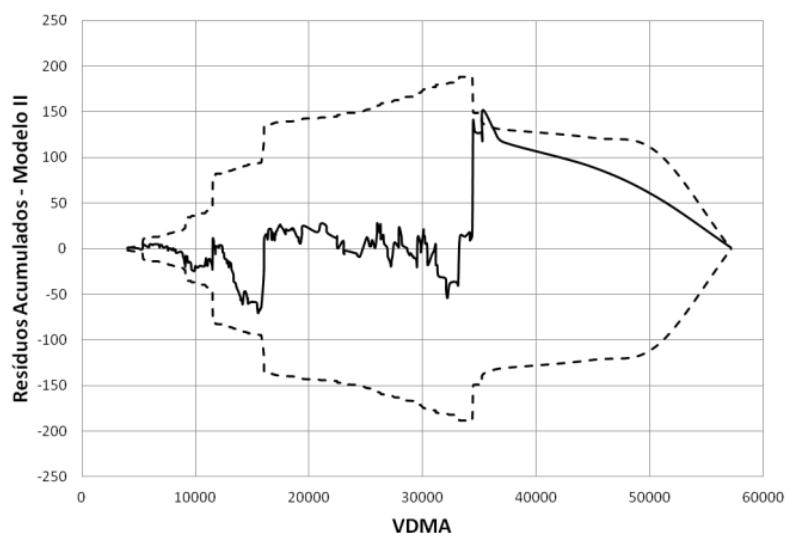
$$\sigma_i^* = \hat{\sigma}_i(n) \sqrt{1 - \frac{\hat{\sigma}_i^2(n)}{\hat{\sigma}^2(N)}} \quad (11)$$

Em que:  $\sigma_i^*$  representa o valor da envoltória para a observação  $i$ ,  $\hat{\sigma}_i(n)$  é o resíduo acumulado para  $i$ ,  $\hat{\sigma}_i^2(n)$  é a variância acumulada para  $i$ ,  $\hat{\sigma}^2(N)$  é o somatório da variância da amostra.

### 2.3.2 Métodos para modelagem: abordagem regime de fluxo veicular macroscópico

A abordagem por RFV macroscópico, apesar do menor nível de produção acadêmica em comparação à entidade-anual, já foi explorada a certo ponto de permitir o desenvolvimento de ferramentas que aperfeiçoem a qualidade dos resultados. Dentre elas, surgem dois principais pontos – detalhados nas próximas seções – que devem ser considerados antes do processo de calibração e validação dos modelos: (i) estimativa do horário mais provável da ocorrência dos acidentes; (ii) agrupamento dos regimes de fluxo veicular similares.

Figura 4 – Exemplo de CURE plot



Fonte: Sobreira e Cunto (2014).

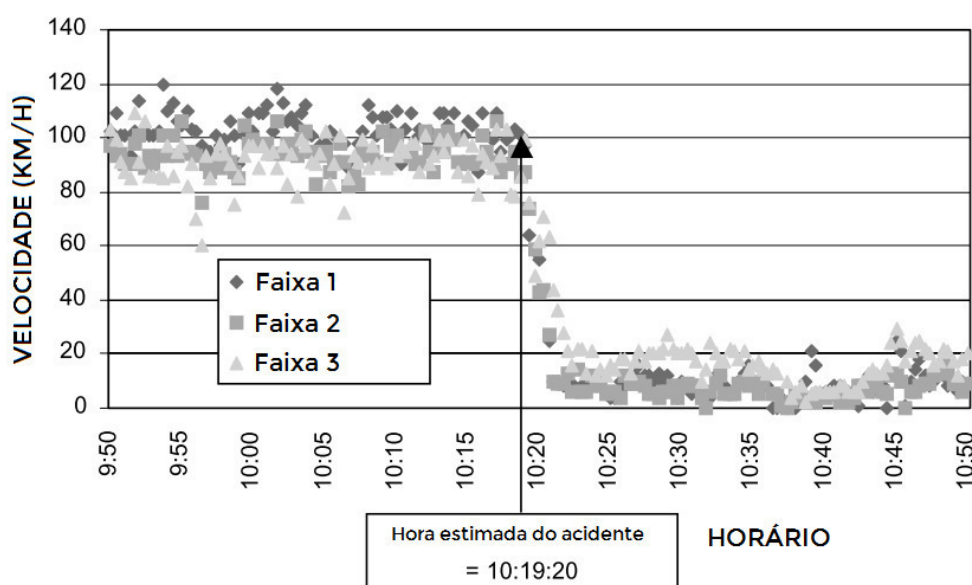
#### 2.3.2.1 Estimativa do horário mais provável de ocorrência do acidente em função de perturbações na condição do fluxo veicular

Ao mesmo tempo em que a utilização de dados desagregados permite uma modelagem mais robusta, manifesta-se a necessidade de uma maior precisão no horário da ocorrência dos acidentes. O emprego dos horários reportados pelos agentes responsáveis pelo atendimento do acidente – que podem ser imprecisos – pode não corresponder às características de fluxo veicular consideradas como precursoras ao acidente, introduzindo relações de causa e efeito errôneas (ROSHANDEL; ZHENG; WASHINGTON, 2015; ZHENG, 2012). Grande parte dos estudos desenvolvidos no tema não considerou a estimativa do horário mais provável de ocorrência dos acidentes (CEDER; LIVNEH, 1982; CHRISTOFOROU; COHEN;

KARLAFTIS, 2011; GOLOB; RECKER; PAVLIS, 2008; IMPRIALOU et al., 2016; LEE; ABDEL-ATY; HSIA, 2006; PIRDAVANI et al., 2015; SOLOMON, 1964; STEMPFEL et al., 2016).

Por outro lado, alguns autores incorporaram técnicas de detecção de acidentes em suas análises. Lee *et al.* (2002), em seu estudo de precursores de acidentes em rodovias, estimaram o horário mais provável de ocorrência dos acidentes a partir de informações de velocidade agregadas em um minuto do laço detector à montante do georreferenciamento da ocorrência. O critério de identificação visual considerou quedas abruptas na velocidade detectada no laço à montante, assumindo que a redução da velocidade foi gerada pelo acidente, e não o contrário. A Figura 5 apresenta o método utilizado pelos autores. Para o acidente apresentado, o horário reportado foi de 10h21m22s, 162 segundos após o estimado pela redução da velocidade. A mesma técnica também foi aplicada por Lee *et al.* (2003) e Theofilatos *et al.* (2018).

Figura 5 – Método de estimação do horário da ocorrência do acidente baseado em dados de velocidade de um laço à montante do acontecimento



Fonte: Adaptado de Lee *et al.* (2002)

Zheng *et al.* (2010) utilizaram dados agregados em intervalos de 20 segundos dos laços detectores à montante e à jusante da localização do acidente para detecção do horário mais provável para avaliar a relação entre oscilações de tráfego e a frequência de acidentes em *freeways*. A identificação visual foi baseada em reduções da velocidade do laço à montante e em incrementos na velocidade e contração no volume do laço à jusante. Esse método – com

variações na agregação temporal dos dados – é o mais comum nos estudos aplicados em vias de fluxo ininterrupto que dispõem de contínuos laços detectores (Abdel-Aty *et al.*, 2005; Lee *et al.*, 2003; Pande e Abdel-Aty, 2006; Zheng, 2012).

Hojati *et al.* (2014) apresentaram um estudo que avalia o tempo da duração total de um acidente incluindo a ocorrência do acidente e o retorno a RFV típicos (*recovery time*) em 14 km de *freeways* em Queensland-AUS a partir de indicadores de velocidade e volume agregados em intervalos de cinco minutos. Os autores utilizaram dados de velocidade também agregados em cinco minutos do laço detector à montante para reconhecimento do horário preciso do acidente. As velocidades médias no período do acidente foram comparadas a valores típicos na via, e; caso reduções substanciais (não quantificadas pelos autores) fossem visualmente identificadas em quatro intervalos consecutivos, o horário era reconhecido.

Outra técnica utilizada na identificação de anomalias na corrente veicular é a Transformada *Wavelet* (TW). A TW é composta por funções que são capazes de decompor séries históricas de dados (volume veicular, velocidade) possibilitando a análise em diversas escalas e níveis temporais. A ferramenta foi amplamente utilizada em estudos voltados para o ambiente rodoviário (JIANG; ADELI, 2004; KIM; REDDY; VANNUCCI, 2004; RIBEIRO; CUNHA, 2017; SAMANT; ADELI, 2000).

Outra frente aplicou atrasos fixos nos horários reportados na tentativa de obtenção das condições precursoras aos acidentes. Christoforou *et al.* (2011), por exemplo, considerou um atraso fixo de seis minutos no horário do acidente para extração das características de tráfego, ou seja, para um acidente reportado às 9h00m, os autores consideraram as condições na via a partir das 8h54m. Outros pesquisadores adotaram o mesmo método com alteração no atraso fixo: (i) Golob *et al.* (2008): 12 minutos; (ii) Quddus *et al.* (2010): 30 minutos.

#### 2.3.2.2 *Técnicas de agrupamentos de dados para definição de agrupamentos de regime de fluxo veicular*

A utilização de dados desagregados em intervalos horários ou sub-horários – além da necessidade de uma melhor estimativa do horário de ocorrência do acidente – também demanda que as variáveis explicativas sejam agrupadas por similaridade, visto que a considerável quantidade de observações por entidade dos regimes de fluxo veicular associada ao caráter raro da ocorrência dos acidentes exige que os dados sejam tratados de forma diferente da tradicional (IMPRIALOU *et al.*, 2016). Por exemplo, supõe-se a análise de cinco interseções semaforizadas que apresentaram frequência de dez acidentes em determinado ano

a partir de dados agregados em intervalos horários. O estudo seria composto por 43.800 intervalos ( $365 \times 24 \times 5$ ) com diversas configurações de características de tráfego, tornando inviável que apenas 50 intervalos com ocorrência de acidente representem a probabilidade de frequência desses.

Com isso, Imprialou *et al.* (2016) sugerem que estudos desagregados sejam conduzidos aplicando agrupamentos de regimes de fluxo veicular (*condition-based*). Os agrupamentos consistem na segmentação de todas as possíveis variáveis presentes nas entidades durante o período estudado, seja ele com ou sem a ocorrência de acidentes. Os autores agruparam cenários a partir de percentis da velocidade (50 percentis) e do volume (quatro percentis) e da categorização das variáveis de geometria, compondo ao todo 2400 cenários. Utilizando a informação do horário reportado de 9.310 acidentes, cada ocorrência foi associada a um cenário. Seguindo com o exemplo apresentado no parágrafo anterior, os 43.800 intervalos horários deveriam ser agrupados em N elementos a partir da similaridade entre eles, com a frequência de acidentes sendo atribuída a cada elemento, facilitando a obtenção de resultados mais confiáveis.

Seguindo o intuito de agregar características similares, diversas técnicas de agrupamentos de dados vêm sendo utilizadas nas últimas décadas principalmente em estudos de engenharia de tráfego com foco na definição de padrões de fluxo veicular e previsão a longo e curto prazo de volumes veiculares (CELIKOGLU, 2013; CHUNG, 2003; ERMAN; ARLITT, 2006; GARCÍA-RÓDENAS; LÓPEZ-GARCÍA; SÁNCHEZ-RICO, 2017; MONTAZERI-GH; FOTOUHI, 2011; SILGU; ÇELIKOĞLU, 2014; ZHENG *et al.*, 2011). Há também alguns casos de pesquisas na área de segurança viária que fizeram uso de técnicas de agrupamento para definição de condições similares de fluxo veicular para associação com as frequências de acidentes (GITELMAN; DOVEH; BEKHOR, 2017; HAMZEHEI; CHUNG; MISKA, 2013; IMPRIALOU *et al.*, 2016; ZHENG, 2012).

O principal conceito do agrupamento de dados é unir elementos de características similares. O algoritmo *K-means* (MACQUEEN, 1967) é uma das técnicas de aglomeração de dados mais utilizada nas pesquisas de engenharia. O *K-means* tem o objetivo de particionar a amostra de N elementos em K agrupamentos com base nas distâncias multidimensionais (normalizadas ou não) entre as variáveis. O algoritmo inicialmente define aleatoriamente valores para os K centroides que, durante o processo iterativo, são alterados buscando a combinação que resulte no menor somatório entre as distâncias entre agrupamentos e seus elementos. O procedimento exige que o número K de agrupamentos seja informado



previamente, definição que é fundamentada na redução do somatório das distâncias entre elementos e centroides com o aumento de K.

Hamzehei *et al.* (2013) aplicaram o *K-means* para definir regimes de tráfego baseados em perfis de velocidades em rodovias para relacionamento com a frequência de acidentes. Os autores testaram a combinações de 2 a 30 agrupamentos, sendo a configuração com  $K = 11$  a que apresentou a conformação mais eficiente, com base no índice de Dunn (*Dunn index*) que mede a relação entre os diâmetros máximos dos agrupamentos e as distâncias entre centroides dos grupos. Os onze agrupamentos foram classificados em cinco tipos de regimes: (i) tráfego livre; (ii) transição tráfego livre-congestionamento; (iii) transição congestionamento-tráfego livre; (iv) congestionamento; (v) tráfego instável.

Chen *et al.* (2014) desenvolveram um estudo de identificação de padrões de tráfego a partir de informações do tempo de viagem. A técnica *K-means* foi utilizada, e o número ótimo de agrupamentos foi definido com uso do indicador da soma dos erros quadráticos entre agrupamentos (SEQA) (*total within-cluster sum of square*) (Equação 12). O número de agrupamentos escolhido pelos autores foi aquele em que a conformação seguinte apresentara redução inferior a 5% na somatória dos desvios.

$$SEQA = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad (12)$$

Em que:  $k$  é o número de agrupamentos;  $x_i$  é um ponto pertencente ao agrupamento  $C_k$ ;  $\mu_k$  é centroide associado ao agrupamento  $k$ .

A técnica *K-means* – conhecida por sua efetividade e simplicidade – apresenta algumas limitações, tais como (i) devido à escolha aleatória das posições iniciais dos agrupamentos, uma má inicialização pode gerar em agrupamentos ineficazes; (ii) a definição prévia do número K de agrupamentos pode causar a junção de grupos que deveriam ser separados ou a separação de conjuntos similares (LINDEN, 2009).

Outro modo de agrupamento de dados é o Agrupamento Hierárquico (AH) (*hierarchical clustering*) que – diferentemente do *K-means* – não necessita de uma prévia informação do número K de agrupamentos. O AH cria uma hierarquia entre cada elemento do banco de dados com base na distância entre eles de duas formas: (i) aglomerativa: inicialmente cada elemento representa um agrupamento, e em cada passo do algoritmo um par de elementos é unido até que exista apenas um grupo; (ii) divisiva: um único agrupamento é

formado por todos os elementos, ocorrendo a separação – a cada passo do algoritmo – dos pares mais heterogêneos até que o número de agrupamentos seja igual ao de elementos.

Weijermars e van Berkum (2005) utilizaram a técnica AH na definição de padrões de fluxo veicular entre dias do ano aplicando dados de volume agregados em intervalos de 15 minutos. Com o emprego do indicador SEQA, os autores inicialmente definiram três tipos de características de agrupamentos: (i) finais de semana, feriados, e dias em época de férias escolares; (ii) dias úteis; (iii) domingos. Em seguida, uma análise segregada por dias úteis e não-úteis permitiu a divisão desses em cinco e quatro agrupamentos, respectivamente.

O método de AH destaca-se do *K-means* – principalmente – por não necessitar da informação inicial do número de conjuntos e por não ser influenciado pelo caráter estocástico inerente às posições iniciais dos agrupamentos. Além disso, o AH permite a visualização dos dados em forma de dendograma, possibilitando um melhor entendimento do sequenciamento de agrupamento/divisão dos conjuntos. Entretanto, devido à maior complexidade do algoritmo e à necessidade do cálculo das distâncias entre elementos a cada iteração, o AH demanda maiores tempos de processamento e capacidade de memória dos computadores (LINDEN, 2009).

Outras técnicas de agrupamento com princípios similares às apresentadas anteriormente, como *DBSCAN* e *AutoClass*, também são encontradas em estudos de engenharia de tráfego (ERMAN; ARLITT, 2006; THANG; KIM, 2011). É importante destacar que em sua maioria os métodos de agrupamento produzem conjuntos com diferentes quantidades de elementos, ou seja, durante a calibração dos modelos uma variável que represente a exposição de cada agrupamento deve ser considerada para que a desigual exposição de certos agrupamentos não influencie na análise dos acidentes.

Além do indicador SEQA, outras técnicas são comumente empregadas para determinação do número ótimo de agrupamentos a ser adotado. Uma delas é o teste da silhueta (*Silhouette test*) que mede, em termos globais, o quão bem cada elemento está associado ao seu agrupamento ao comparar a distância entre o centroide de cada elemento e o centroide de seu agrupamento com a distância entre o centroide de cada elemento e o centroide do agrupamento mais próximo. Outro teste é o *Gap Statistic* (TIBSHIRANI; WALTHER; HASTIE, 2001) que compara a soma dos erros entre agrupamentos da amostra utilizada com o somatório dos erros caso fosse aplicada uma amostra com distribuição aleatória (sem estrutura de agrupamentos).

### 2.3.2.3 Técnicas de calibração e validação da modelagem de regime de fluxo veicular macroscópico

Levando em consideração a maior complexidade na característica do banco de dados da modelagem dos regimes de fluxo veicular macroscópicos, que possuem mais detalhamento nas informações por cada entidade e que podem apresentar diversas configurações; foi verificado que os estudos, em geral, seguem três frentes para calibração dos resultados: (i) associação dos regimes à frequência de acidentes (GARBER; WU, 2001; IMPRIALOU et al., 2016; LORD; MANAR; VIZIOLI, 2005; MARTIN, 2002; WANG; QUDDUS; ISON, 2009); (ii) associação dos regimes à probabilidade de ocorrência dos acidentes (PEI; WONG; SZE, 2012; THEOFILATOS et al., 2018; ZHENG, 2012); (iii) não realizaram calibração de modelos, apresentando os dados apenas via análise exploratória (HAMZEHEI; CHUNG; MISKA, 2013; STEMPFEL et al., 2016).

Na linha dos estudos que consideram a frequência de acidentes com medida de desempenho da calibração, Garber e Wu (2001) aplicaram modelos de Poisson, BN, PIZ e BNIZ em um banco de dados em que 60% dos RFV considerados exibiu frequência de acidentes igual a zero. Os modelos que apresentaram melhores resultados foram o BN e o PIZ, com o primeiro correspondendo a maiores valores do logaritmo da verossimilhança e o segundo, menores erros globais. Para avaliar a relação entre 2400 regimes de fluxo veicular e 9.310 acidentes, Imprialou et al. (2016) utilizaram modelos multivariados com distribuição Poisson log-normal. A escolha dos modelos mais ajustados foi realizada com emprego do *Deviance Information Criteria (DIC)*, um indicador aplicado em modelos bayesianos com interpretação similar ao *AIC*.

Zheng (2012) empregou modelos logísticos na tentativa de associar diferentes regimes de fluxo veicular a probabilidades de ocorrência de acidentes. O autor apenas avaliou a significância de cada variável e não utilizou nenhuma técnica para verificar o ajuste do modelo calibrado. Theofilatos et al. (2018) aplicaram uma extensão dos modelos logísticos que considera correção no viés da amostra de eventos raros e estimação dos parâmetros a partir da técnica de penalização da máxima verossimilhança na comparação entre regimes de fluxo veicular com e sem a ocorrência de acidentes. A comparação entre modelos foi realizada com utilização do indicador *McFadden's R<sup>2</sup>* que mede a razão entre o logaritmo da verossimilhança do modelo com as variáveis e do modelo nulo.

## 2.4 Considerações finais

A revisão bibliográfica apresentou que os estudos de segurança viária aplicados em interseções semaforizadas mostraram que variáveis de exposição (VDMA) e de geometria e operação das vias (número de faixas, número de conversões à esquerda e à direita) são as que estão mais relacionadas à frequência de acidentes de trânsito. Entretanto, a grande maioria dessas pesquisas utilizou dados com agregações anuais, o que dificulta a inferência sobre a real influência de variáveis de tráfego que oscilam dentro de um mesmo dia (fluxo veicular, velocidade, grau de saturação, entre outras). A utilização de dados de tráfego desagregados (agregação sub-horária) em estudos de vias de fluxo ininterrupto mostrou que, em geral, condições de fluxo de congestionamento e de transição foram aquelas mais associadas à frequência de acidentes. A lacuna na literatura envolvendo a segurança viária de ambientes altamente adensados de interseções semaforizadas e variáveis de tráfego desagregadas ficou clara com o desenvolvido do capítulo.

A revisão mostrou, ainda, a importância de estimativas mais precisas para o horário dos acidentes, além da necessidade de agrupamento das condições de fluxo veicular com o uso de extensos bancos de dados.

### 2.4.1 Hipóteses a serem verificadas no trabalho

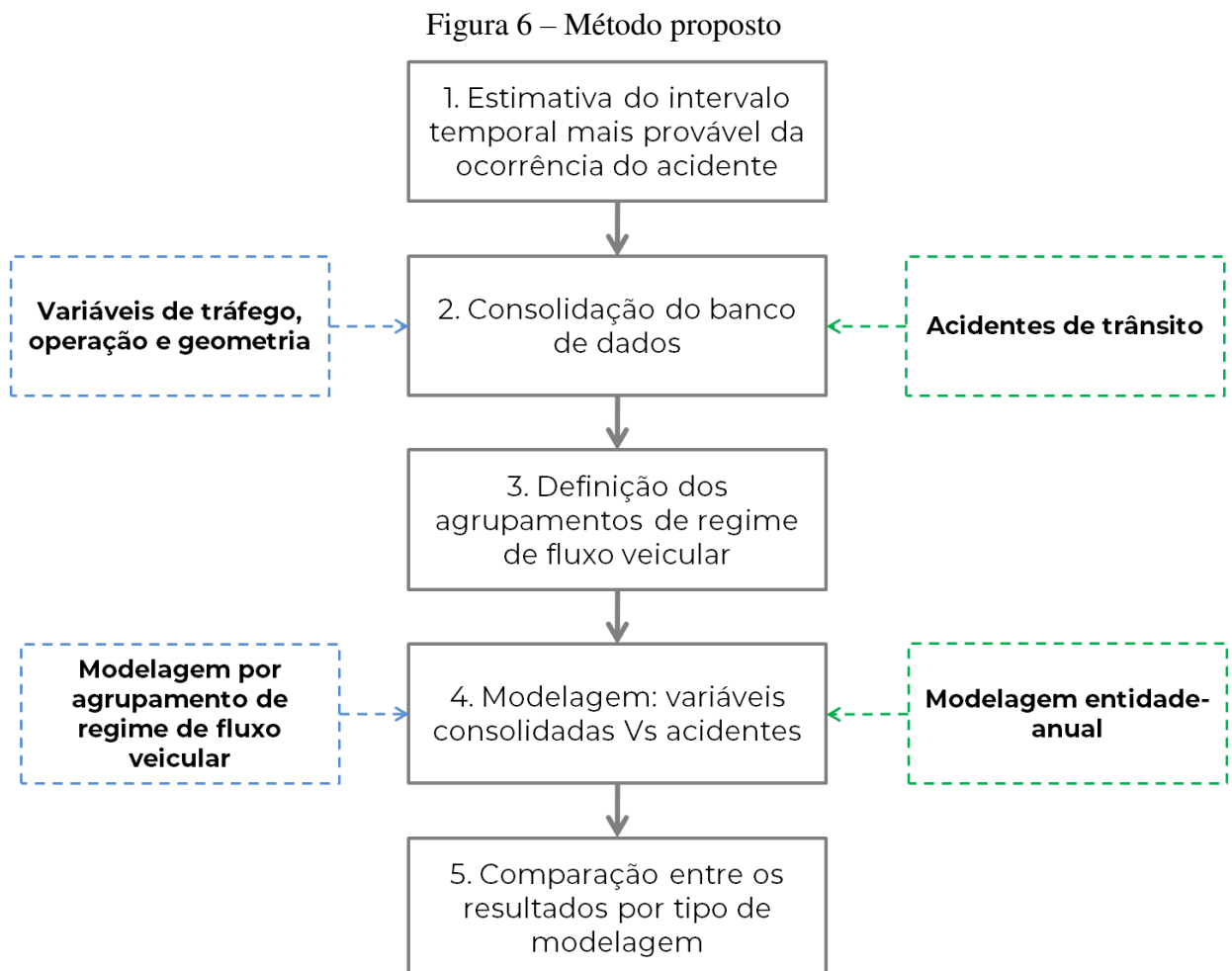
Apesar de a revisão bibliográfica não apresentar estudos com dados desagregados realizados para o meio urbano adensado de interseções semaforizadas, levantaram-se hipóteses sobre a relação entre a frequência e severidade dos acidentes e as características dos RFV – considerando os achados para o ambiente ininterrupto e conclusões de estudos aplicando dados agregados anualmente para locais semaforizados. As hipóteses, visando guiar a discussão na seção dos resultados, são:

- 1) RFV congestionados (altos *fluxos veiculares* e baixas *velocidades*) estão associados a maiores frequências de acidentes, principalmente os de baixa severidade;
- 2) RFV não congestionados (baixos-médios *fluxos veiculares* e médias-baixas *velocidades*) estão associados a menores frequências de acidentes, mas com predominância de ocorrências severas;
- 3) Colisões traseiras são mais frequentes em RFV de alta demanda veicular;

- 4) O aumento no *número de faixas* está associado a maiores frequências de acidentes (CUNTO; CASTRO NETO; BARREIRA, 2011);
- 5) O aumento no *número de aproximações* está relacionado a menores frequências de acidentes (XIE et al., 2013);
- 6) O aumento no número de ciclos semafóricos por hora (associado a maiores repetições do movimento parada-arranque) está relacionado a maiores frequências de acidentes.

### 3 MÉTODO

O método proposto (Figura 6) é composto por cinco macro etapas: (i) estimativa do intervalo temporal mais provável da ocorrência do acidente a partir da identificação de perturbações na corrente de tráfego; (ii) consolidação de bancos de dados com informações de tráfego, de uso do solo da operação e geometria das vias e dos acidentes de trânsito; (iii) definição dos agrupamentos de regimes de fluxo veicular (ARFV); (iv) modelagem das variáveis consolidadas e dos acidentes de trânsito considerando ARFV e entidades; (v) comparação entre as variáveis, os parâmetros, os ajustes e as inferências sobre o fenômeno obtidos com as modelagens por agrupamentos e por entidades. As próximas seções apresentam o detalhamento de cada uma das etapas.

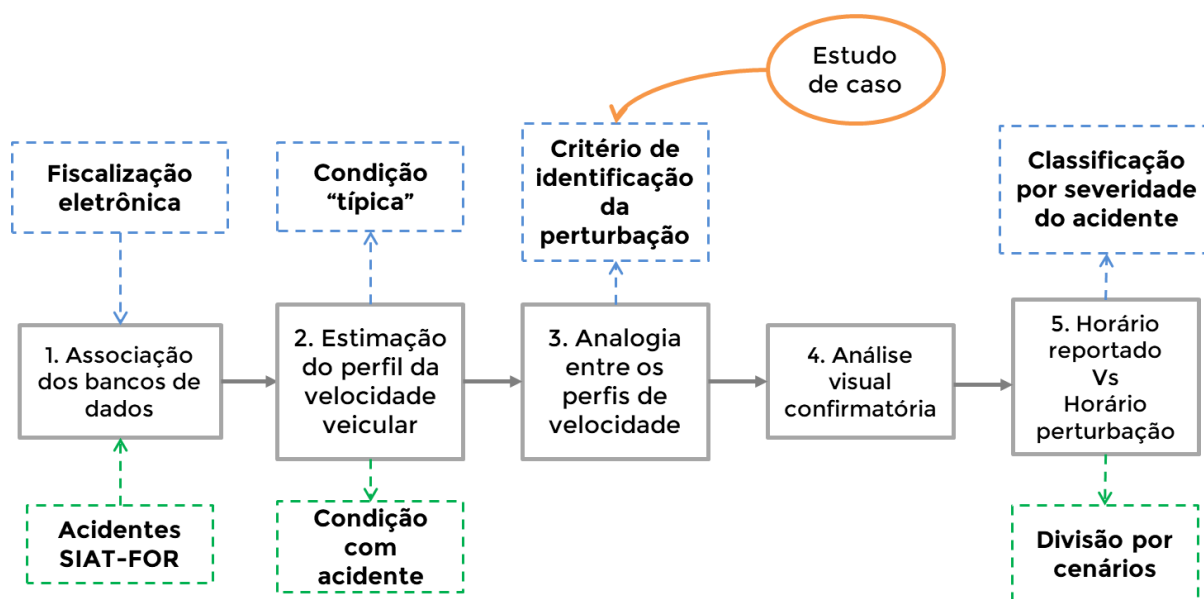


Fonte: Elaborada pelo autor

### 3.1 Estimação do horário mais provável da ocorrência do acidente a partir de perturbações na velocidade da corrente de tráfego

A utilização de RFV agregados em intervalos horários ou sub-horários demanda uma maior precisão na estimativa do horário de ocorrência dos acidentes, visto que os registros realizados pelos agentes de trânsito podem ser imprecisos, fazendo com que relações de causa e efeito errôneas sejam atribuídas na associação dos acidentes às suas condições precursoras. Na Figura 7, apresenta-se o método proposto para estimação do horário mais provável da ocorrência do acidente de trânsito a partir de perturbações na velocidade da corrente de tráfego que permitam o desenvolvimento de uma comparação com os horários reportados no banco de dados de acidentes.

Figura 7 – Método proposto para análise entre tempo reportado da ocorrência de acidentes e perturbações na condição na velocidade em interseções semaforizadas



Fonte: Elaborada pelo autor

A primeira etapa do método consistiu na associação de dois bancos de dados: (i) acidentes de trânsito e (ii) fiscalização eletrônica. A Prefeitura de Fortaleza armazena os dados históricos de acidentes de trânsito em uma plataforma denominada Sistema de Informações de Acidentes de Trânsito da cidade de Fortaleza (SIAT-FOR) desde o ano de 2004. O SIAT-FOR recebe informações de acidentes de diversas fontes (agentes da Prefeitura, SAMU, Perícia Forense, hospitais, IML, entre outras), as quais são tratadas e georreferenciadas pelos gestores da Prefeitura. O banco de dados fornece importantes

características das ocorrências, como: data, horário (hora e minuto), severidade (danos materiais, feridos ou fatais), tipo do acidente (colisão traseira, colisão transversal, atropelamentos, entre outros) e tipo de veículos/usuários envolvidos.

Os dados de fiscalização eletrônica são disponibilizados em arquivos texto estruturados em passagens individuais veiculares, ou seja, cada linha de observação representa uma passagem. Para cada observação as seguintes informações são concedidas: código do equipamento, data da passagem, horário da passagem (hora, minuto e segundo), faixa da passagem, velocidade permitida na via, velocidade medida pelo equipamento, comprimento medido do veículo e classificação do veículo (motocicleta, carro, caminhão, ônibus ou não identificado).

A associação entre os dois bancos de dados teve o intuito de associar cada equipamento de fiscalização a determinados acidentes que ocorreram em suas redondezas. Devido à imprecisão do processo de georreferenciamento dos acidentes ocasionada pela técnica de coleta da localização das ocorrências pelos agentes de trânsito, que utiliza a numeração do lote lindeiro ou nome das vias do cruzamento, decidiu-se por implementar algumas medidas para que um maior controle da posição dos acidentes fosse possível: (i) seleção equipamentos situados em interseções semaforizadas; (ii) coleta de acidentes que ocorreram em um raio de 30 metros dos equipamentos.

Associados os horários reportados dos acidentes aos equipamentos de fiscalização eletrônica, determinaram-se os perfis da velocidade veicular em condições “típicas” (sem ocorrência de acidente) e em condições com ocorrência de acidente. Devido à incerteza da precisão do horário reportado no banco de dados, decidiu-se avaliar um intervalo de três horas em torno do mesmo (hora do acidente, hora anterior e hora seguinte).

Para definir a condição “típica” no fluxo veicular, cinco dias (uma e duas semanas antes e uma, duas e três semanas após a data do acidente) foram avaliados considerando o mesmo intervalo temporal de três horas e o mesmo dia da semana. Foram determinadas as velocidades médias do intervalo de três horas para os cinco dias, em que o dia que apresentou o valor da mediana entre as médias foi selecionado como típico. Por exemplo, para um acidente que foi reportado às 14h30m no dia 15/05/2017 foram extraídas as velocidades médias para o intervalo de 13h00m a 15h59m das datas 01/05/2017, 08/05/2017, 22/05/2017, 29/05/2017 e 05/06/2017, sendo o dia considerado como típico aquele que exibiu a mediana entre as velocidades médias obtidas.

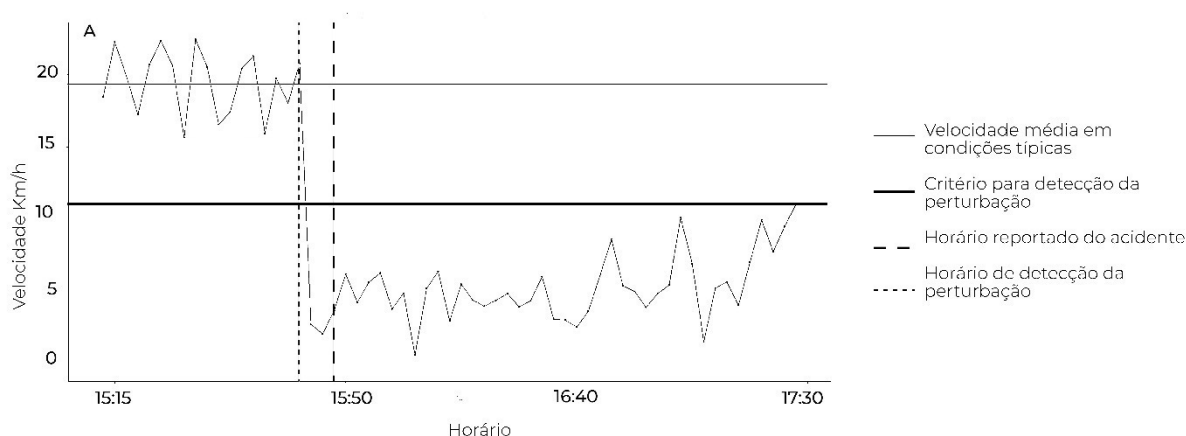
Com o intuito de automatizar a detecção de perturbações na velocidade, foi necessária a definição de um critério que possibilite o reconhecimento de oscilações na mesma. Um



estudo piloto foi desenvolvido para determinação do critério de maior detecção correta das perturbações na velocidade. Foram aplicados três parâmetros com base na média e no desvio padrão da velocidade em condições típicas e no intervalo de agregação da velocidade: caso a velocidade agregada em X minutos permaneça abaixo do valor da média subtraída de Y desvios padrão durante Z minutos consecutivos, a perturbação é reconhecida.

Cada parâmetro foi baseado em agregações da velocidade (X) em 1 e 3 minutos com subtração de (Y) 2 e 3 desvios padrão durante (Z) 6 e 9 minutos. A Figura 8 traz um exemplo de aplicação do método em que a velocidade foi agregada em 3 minutos (X), sendo o parâmetro para detecção da perturbação (linha grossa) obtido com redução da velocidade média em condições típicas (linha fina) de 2 desvios padrão (Y). Para que uma perturbação seja detectada pelo critério ela deve permanecer por 6 ou 9 minutos (Z) abaixo da linha grossa. No total, o estudo piloto compreendeu as oito combinações dos três parâmetros.

Figura 8 – Exemplo de aplicação do método para detecção de perturbações na velocidade



Fonte: Elaborada pelo autor

Para os acidentes em que foi possível a identificação correta da oscilação na velocidade, gráficos foram gerados para desempenho de uma análise visual complementar. Nessa análise, observaram-se individualmente os gráficos correspondentes a cada acidente, considerando como válidos aqueles em que foi possível detectar o momento exato do início da perturbação na velocidade, correspondendo a uma mudança brusca no padrão do gráfico e, desta forma, aumentando a confiabilidade no critério utilizado. Essa abordagem é discutida em mais detalhes na seção dos resultados. A melhor combinação de agregação de velocidade média, de desvio padrão e de tempo de duração foi a que produziu a maior porcentagem de detecção automática e de casos visualmente validados.

Por fim, com a aplicação do algoritmo selecionado em todo o banco de dados de acidentes, compararam-se os horários reportados dos acidentes com os horários de detecção das perturbações na velocidade, objetivando uma estimativa média da diferença entre eles. Os resultados obtidos foram segregados por severidade do acidente, por horário de ocorrência do acidente, por tipo de acidente e por quantidade de laços detectores que identificaram a perturbação de um mesmo acidente. As duas últimas segmentações são importantes, pois se supõe que, por exemplo, colisões transversais (controle da localização do acidente) ou ocorrências detectadas por dois ou mais laços (possibilidade de validação do horário da perturbação) resultariam em estimativas mais confiáveis.

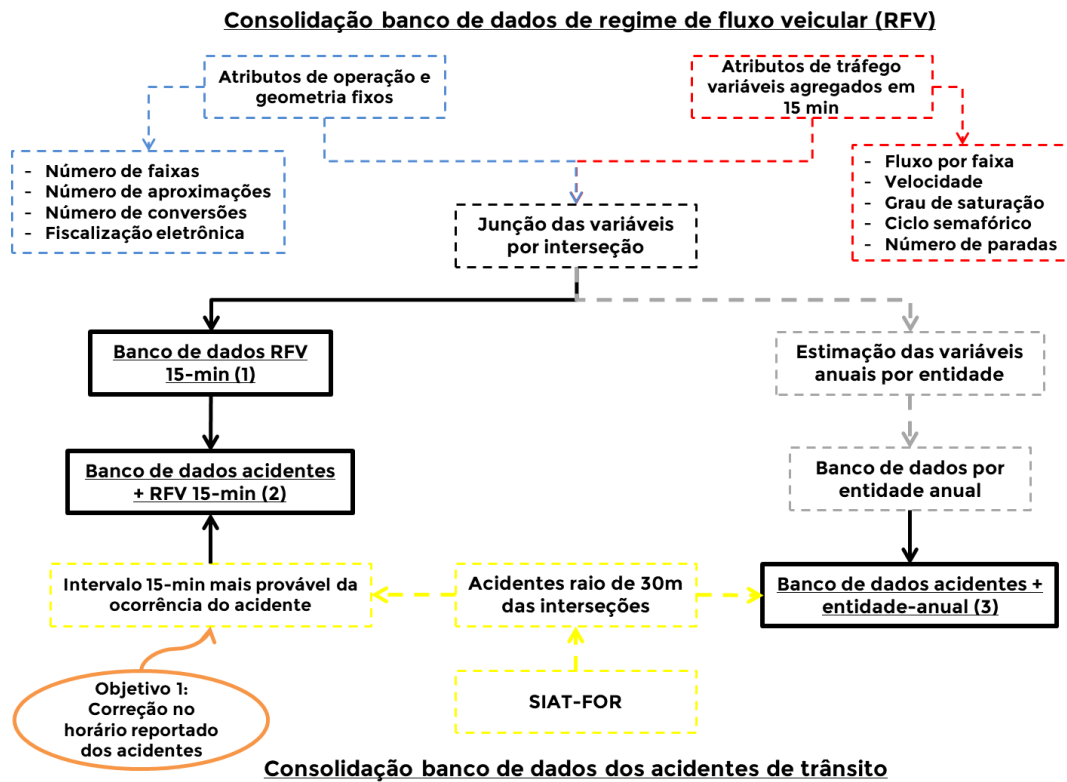
### **3.2 Consolidação do banco de dados de regimes de fluxo veicular e de acidentes**

Para o desenvolvimento das etapas de definição dos agrupamentos e de modelagem estatística, demandou-se a consolidação de três bancos de dados contendo informações de RFV (características de tráfego, geometria e operação das vias) e dos acidentes de trânsito: (i) banco de dados de RFV agregados em intervalos de 15 minutos; (ii) banco de dados dos acidentes associados a um dado RFV em intervalos de 15 minutos; (iii) banco de dados dos acidentes associados a entidades com informações anuais. A Figura 9 apresenta o método proposto.

As variáveis a serem coletadas foram divididas em dois principais grupos: (i) atributos fixos de operação e geometria das interseções que não oscilam durante um mesmo dia; (ii) atributos variáveis de tráfego que podem oscilar dentro de um mesmo dia. As variáveis de geometria e operação das vias de Fortaleza (*número de faixas, número de aproximações semaforicas, número de movimentos de conversão e presença de fiscalização eletrônica*) foram obtidas de bases georreferenciadas fornecidas pela Prefeitura da cidade e de imagens das plataformas *Google Maps* e *Google Earth*.

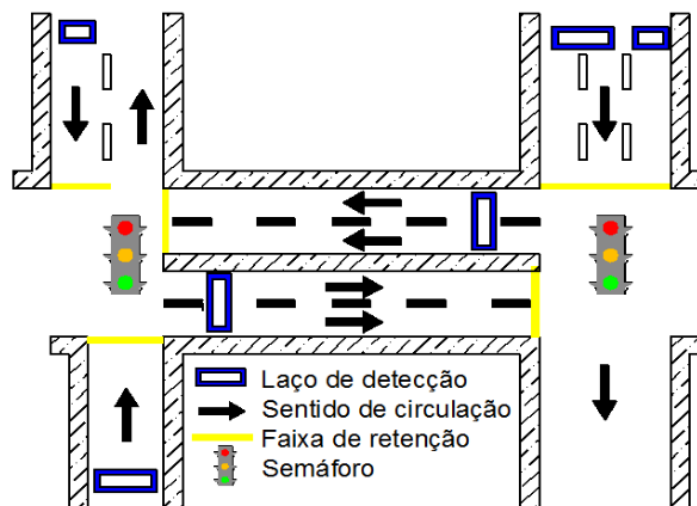
Os atributos variáveis de tráfego foram coletados do sistema de gerenciamento e controle semaforico da cidade denominado de *SCOOT (Split, Cycle and Offset Optimization Technique)*. O sistema do *SCOOT* usualmente é utilizado na otimização dos ciclos semaforicos e na detecção de congestionamentos na malha viária. Sua estrutura é formada por laços de detecção veicular localizados, geralmente, no início do segmento da aproximação semaforica (Figura 10). Assim, para que o sistema esteja em perfeito funcionamento, os laços de todas as aproximações devem estar em atividade.

Figura 9 – Método proposto para consolidação dos bancos de dados de RFV e dos acidentes



Fonte: Elaborada pelo autor

Figura 10 – Ilustração do posicionamento dos laços detectores do *SCOOT*



Fonte: Aquino (2012)

O funcionamento dos laços do *SCOOT* é baseado na ocupação veicular, possibilitando a estimação de variáveis básicas, obtidas pelo sistema, e derivadas, estimadas a partir das

variáveis básicas. Dentre as variáveis básicas destacam-se o *fluxo veicular*, o *número de paradas* veiculares, o *atraso veicular* e o *grau de saturação* da aproximação. No grupo das variáveis derivadas ressaltam-se a *velocidade veicular* (estimada a partir da Equação 13), o *índice de congestionamento* e a *duração da viagem*. É importante destacar que a *velocidade* não é diretamente medida pelos laços e sim estimada com base no tempo de viagem em fluxo livre e o atraso estimado pelo sistema, ou seja, o indicador inclui na *velocidade* o tempo em que os veículos estiveram parados. Todas as variáveis são disponibilizadas em intervalos de 15 minutos. Salienta-se, ainda, que os laços do *SCOOT* – algumas vezes devido à falta de manutenção e outras a falhas habituais do sistema – não apresentam informações contínuas do tráfego, assim, é necessário que os dados passem por filtros antes das aplicações.

$$V = 3,6 * \frac{L}{(t+d)} \quad (13)$$

Em que:  $V$  representa a velocidade média em km/h;  $L$  é o comprimento entre o laço detector e a faixa de retenção em metros;  $t$  é o tempo de cruzeiro no trecho inserido no banco de dados em segundos;  $d$  é o atraso veicular médio estimado pelo *SCOOT* em segundos.

A utilização de dados obtidos por aproximação possibilitou duas abordagens: (i) análise individual por aproximação; (ii) análise combinando as aproximações que compõem a interseção semaforizada. O estudo por aproximação é vantajoso por manter a desagregação das informações de tráfego, visto que a agregação por interseção pode levar a perda de importantes características, caso a composição de fluxo ou velocidade seja significativamente diferente entre as aproximações. Entretanto, a abordagem por interseção traz outros importantes benefícios, principalmente com relação à localização dos acidentes que, por serem usualmente registrados no cruzamento, dificultaria a obtenção de amostras significativas na análise por aproximação. Além disso, a facilidade de obtenção automática das variáveis, a partir da utilização de *buffers* em softwares de análise espacial, e a menor complexidade na definição dos regimes de fluxo veicular, visto que o estudo por aproximação levaria à segregação de via principal e via secundária, aumentando a variedade das combinações de condições, tornaram a abordagem por interseção mais viável para condução da pesquisa.

Definida a abordagem por interseção, surgiu a necessidade de tratamento das informações por aproximação para que seja realizada a agregação por interseção. Determinou-se que, para consideração de uma dada interseção, todas as aproximações deveriam conter

informações de tráfego válidas por, pelo menos, 50% de um ano. A combinação por interseção só ocorreu caso todas as aproximações apresentassem informações para o intervalo de 15 minutos em questão. As variáveis discretas (*fluxo, número de paradas*, entre outras) foram somadas por aproximação e as variáveis contínuas (*velocidade, grau de saturação*) foram consideradas com base na média ponderada por fluxo da aproximação. Por exemplo: uma interseção com duas aproximações A (fluxo = 1000 ve/h e velocidade média = 30 km/h) e B (fluxo B = 2000 ve/h e velocidade média = 15 km/h) apresentou variáveis para o intervalo de 15 minutos de fluxo = 3000 ve/h e velocidade média = 20 km/h.

Unidas as variáveis de tráfego para cada intervalo temporal às características de operação e geometria das interseções, obteve-se um extenso banco de dados (*banco de dados RFV – 15 min*, Figura 9) que foi utilizado na definição dos agrupamentos. A estimação das variáveis anuais para composição do banco de dados por entidades foi realizada com base na média das variáveis contemplando apenas os dias que apresentaram informações válidas para o período das 24 horas do dia.

Os acidentes de trânsito foram coletados a partir de um raio de 50 metros de cada uma das interseções semaforizadas. O *banco de dados acidentes + entidade-anual* foi obtido com a combinação entre o somatório de acidentes por entidade e as variáveis coletadas ou estimadas para cada uma das interseções. Cada linha de observação desse banco de dados representa uma interseção semaforizada com sua frequência de acidentes e suas características de tráfego agregadas anualmente e seus atributos de geometria e de operação da via.

Para consolidação do *banco de dados acidentes + RFV – 15min* foi necessária a aplicação dos resultados obtidos na primeira etapa para estimação do mais provável intervalo temporal de ocorrência do acidente. Como apenas quatro interseções semaforizadas possuem simultaneamente sistema *SCOOT* e fiscalização eletrônica, foi inviável a utilização apenas dos acidentes em que foi possível a identificação do horário mais provável de ocorrência, pois resultaria em uma amostra consideravelmente pequena. Assim, utilizando os resultados obtidos para todas as aproximações semaforizadas avaliadas no objetivo I, decidiu-se trabalhar com uma “correção” média a ser aplicada no horário de reportagem de todos os acidentes, visando obter – na maioria dos casos – informações precursoras à ocorrência dos mesmos. A “correção” média foi obtida com o valor do 95º percentil das diferenças estimadas entre o horário reportado do acidente e o horário da detecção da perturbação na velocidade.

Identificado o intervalo de 15 minutos de cada acidente, combinaram-se essas informações com o *banco de dados RFV – 15 min* para obtenção do banco de dados que durante a etapa de modelagem será utilizado na quantificação do número de acidentes em

cada RFV. Cada observação desse banco de dados descreve um acidente de trânsito com suas informações próprias (horário de ocorrência, tipo, severidade) e suas características de tráfego, de geometria e de operação. Ao longo da exposição dos bancos de dados obtidos, análises exploratórias entre as variáveis dependentes e independentes foram efetivadas, visando ao levantamento de hipóteses sobre a relação entre as mesmas.

### **3.3 Definição dos agrupamentos de regimes de fluxo veicular**

A definição dos agrupamentos de regime de fluxo veicular (ARFV) foi realizada a partir da agregação das informações do *banco de dados RFV 15-min (1)*, cujo método é apresentado na Figura 11. As diferentes configurações de ARFV, produto final desta etapa, foram utilizadas na seção da modelagem. O processo de obtenção dos agrupamentos pode gerar resultados em infinitas dimensões, valor que é função do número de variáveis avaliadas, assim, decidiu-se por realizar esse procedimento utilizando apenas duas variáveis de tráfego, visando facilitar a exposição e o entendimento dos resultados. Essas variáveis foram selecionadas com uso de gráficos de dispersão e de índices de correlação entre as características de tráfego. Os outros atributos de tráfego e as variáveis de geometria e operação foram incorporados aos agrupamentos de forma categórica (a ser detalhado a seguir).

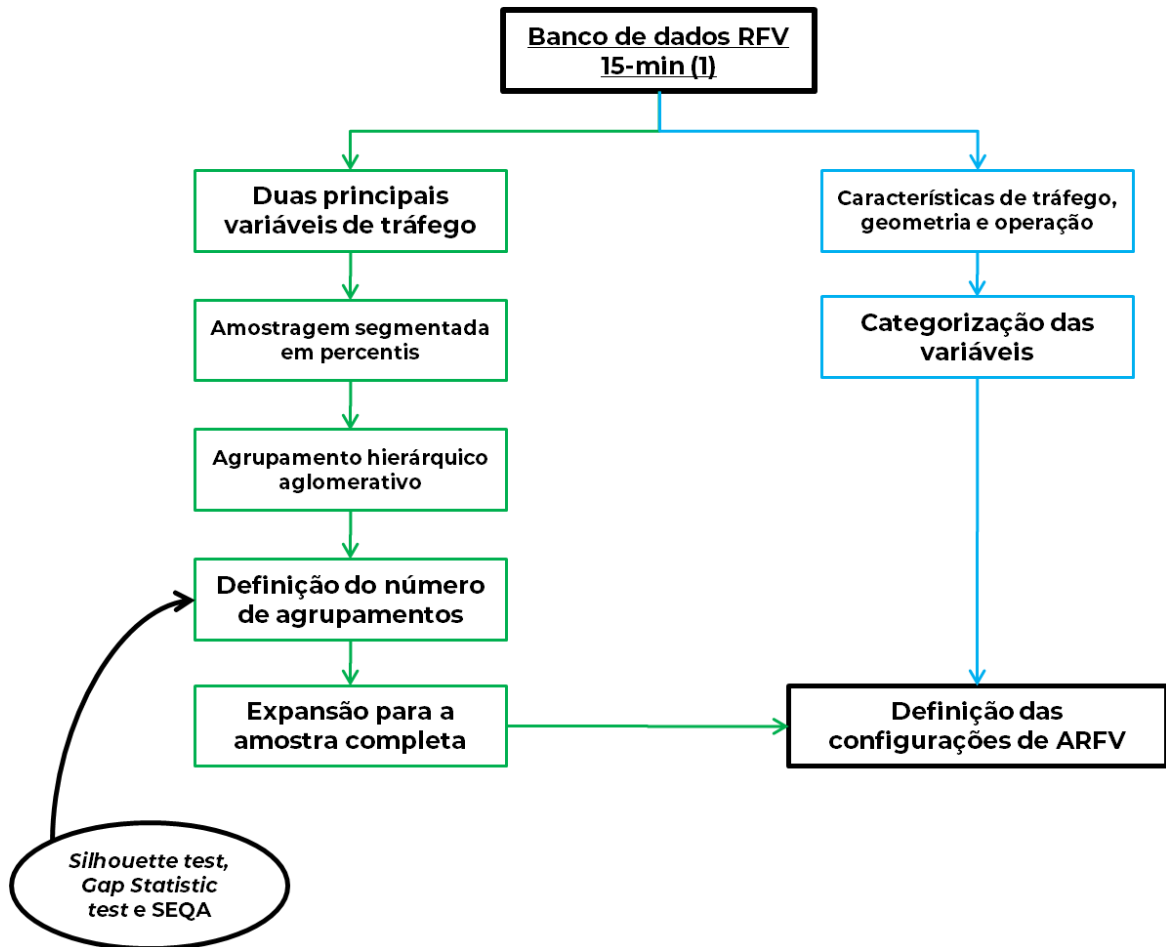
Para definição dos conjuntos homogêneos das duas variáveis de tráfego selecionadas, aplicou-se a técnica de agrupamentos hierárquicos utilizando o método aglomerativo com distâncias quadradas (*ward distance*), que parte de uma configuração em que cada elemento representa um agrupamento agregando-os até que um único agrupamento contendo todos os elementos seja formado. Decidiu-se utilizar a técnica de agrupamentos hierárquicos, em vez da *K-means*, visto que a primeira traz o benefício de não contar com a aleatoriedade no momento da seleção dos pontos iniciais dos agrupamentos, como destacado na seção anterior.

A técnica de agrupamentos hierárquicos utiliza as distâncias normalizadas entre elementos como métrica para geração dos agrupamentos, ou seja, um banco de dados com 500 mil observações demandaria uma matriz distância de dimensões 500.000 x 500.000, sendo inexecutável em computadores de capacidade normal. Assim, devido ao extenso tamanho do *banco de dados RFV 15-min*, requereu-se que um processo de amostragem fosse realizado para viabilizar a execução do algoritmo da técnica de agrupamentos.

Na tentativa de obter uma amostra reduzida que apresentasse distribuição e amplitude similar à amostra completa, aplicou-se uma técnica de amostragem segmentada com base em

percentis. Por exemplo, para gerar uma amostra reduzida bivariada com 5.000 observações, segmentam-se cada uma das variáveis da amostra completa em 2.500 percentis. Uma aplicação prévia do algoritmo mostrou que amostras com até 10.000 observações são exequíveis.

Figura 11 – Método proposto para definição dos agrupamentos de regimes de fluxo veicular



Fonte: Elaborada pelo autor

A definição do número de agrupamentos ( $K$ ) a ser estudado foi baseada em três testes/indicadores: (i) *Silhouette test*; (ii) *Gap Statistic test*; (iii) soma dos erros quadráticos entre agrupamentos (SEQA). Os dois primeiros indicam um número ótimo  $K$  a ser selecionado. Já a SEQA aponta a redução nas distâncias elemento-centroide com o incremento no número de agrupamentos. O método do cotovelo (*elbow method*) utiliza a SEQA sugerindo que  $K$  seja visualmente escolhido quando for observado que o incremento de  $K$  não traz benefícios significativos ao somatório dos erros; representando, graficamente, o vértice de um cotovelo. Outra abordagem empregando a SEQA é aplicada em Chen *et al.*

(2014), que selecionaram como configuração ótima o maior número K que trouxe benefício aos erros entre agrupamentos de 5% quando comparados aos erros de  $(K - 1)$ .

Associada cada observação da amostra reduzida a um agrupamento, combinaram-se todos os elementos da amostra completa a um agrupamento com base na menor distância normalizada quadrada entre elemento-centroide.

Os outros atributos de tráfego, que não foram considerados nos agrupamentos acima, e as características de geometria e operação das interseções foram incorporados aos agrupamentos de forma categórica, as quais foram definidas com base nos quartis das variáveis. A decisão de inclusão ou não de uma variável foi apoiada na correlação entre as mesmas, ou seja, atributos altamente relacionados não foram incorporados simultaneamente em uma mesma configuração. Um exemplo: supondo que a conformação ótima dos agrupamentos das variáveis de tráfego principais foi de dez agrupamentos e que se deseja incluir as variáveis *número de faixas*, segmentada em três categorias e *grau de saturação*, segmentada em cinco categorias. O número de agrupamentos final dessa configuração é de 150 agrupamentos  $(10*5*3)$ , caso exista pelo menos um elemento para cada combinação. Diversas configurações de RFV foram desenvolvidas para aplicação da modelagem estatística.

Após a definição dos ARFV, realizou-se uma análise exploratória entre as variáveis pertencentes aos agrupamentos e os acidentes de trânsito, apresentando os gráficos com as frequências de acidentes com danos materiais e com vítimas por tipo de RFV e as matrizes de correlação entre as variáveis.

### **3.4 Modelagem da relação entre agrupamentos de regimes de fluxo veicular e os acidentes e comparação entre os resultados obtidos com a modelagem por entidades e por agrupamentos**

Nesta seção são apresentados os métodos (Figura 12) para alcançar os objetivos específicos III e IV contemplando as modelagens entre as variáveis consolidadas e os acidentes em abordagens por ARFV e por entidades, além da comparação entre os resultados obtidos por tipo de modelagem. As análises foram realizadas segregando os acidentes por severidade: (i) acidentes apenas com danos materiais; (ii) acidentes com vítimas feridas ou fatais.

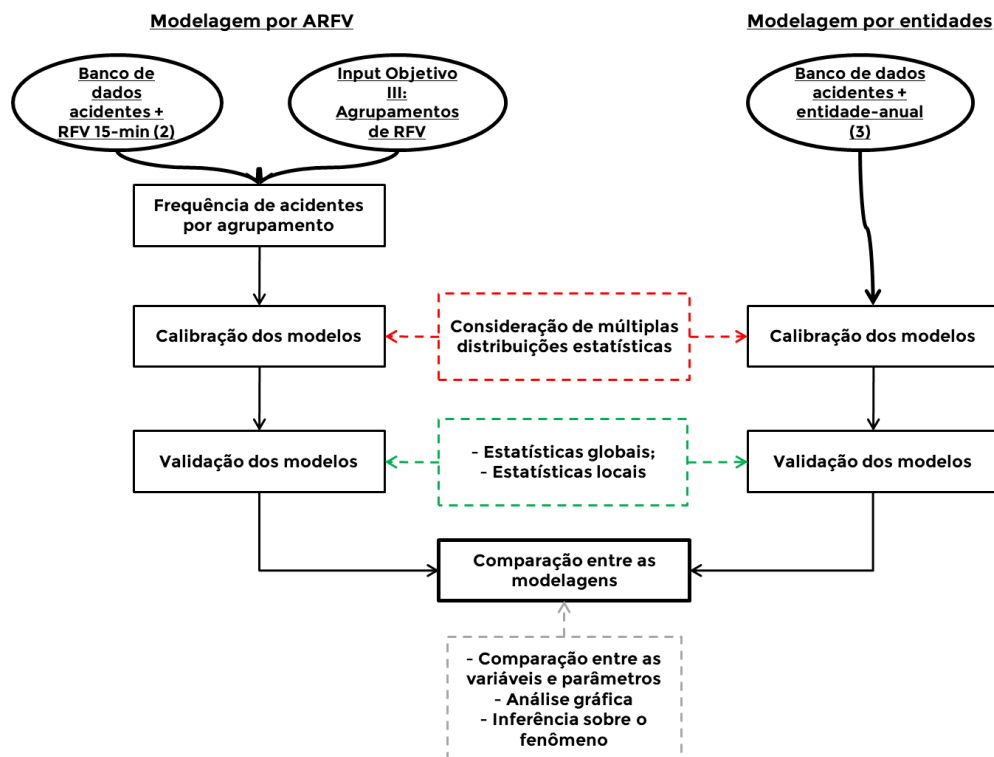
A modelagem por ARFV utilizou as configurações de agrupamentos definidas na seção anterior, associando-os ao *banco de dados acidentes + RFV 15-min (2)*, permitindo a



obtenção da frequência de acidentes para cada agrupamento. Como a quantidade de intervalos de 15 minutos varia para cada agrupamento, decidiu-se utilizar como variável dependente a frequência de acidentes em função da exposição, que foi incorporada ao modelo como *offset*. Ou seja, para cada agrupamento foi estimada a frequência esperada de acidentes em 15 minutos.

Inicialmente, os modelos foram calibrados considerando as clássicas distribuições de Poisson e Binomial Negativa. Para definição da distribuição mais adequada, o indicador de ajuste à distribuição estatística ( $\sigma_d$ ), que compara a dispersão existente nos dados de acidentes com a dispersão considerada pela distribuição estatística, foi utilizado. Valores próximos a um são indicativos de bom ajuste dos dados à distribuição, já valores inferiores a um sugerem que os dados são menos dispersos que a distribuição e valores superiores a um significam que os dados são mais dispersos que a distribuição escolhida. Caso nenhuma das duas distribuições fosse apropriada para a modelagem dos ARFV, outras possibilidades, como modelos inflados de zero, seriam aplicadas; entretanto os modelos de Poisson e Binomial Negativo se mostraram adequados para uma modelagem satisfatória da frequência de acidentes neste trabalho.

Figura 12 – Método proposto para modelagem por agrupamento e por entidade e para comparação dos resultados obtidos por tipo de abordagem



A principal diferença entre essa modelagem e as modelagens tradicionais é que a inclusão ou retirada de uma variável altera a quantidade de agrupamentos, não permitindo que técnicas de seleção dos melhores modelos, como *forward* ou *stepwise*, sejam naturalmente utilizadas. Além disso, os parâmetros AIC e logaritmo da máxima verossimilhança se tornam incomparáveis entre os modelos, pois cada um contém um número diferente de observações. Assim, propôs-se uma adequação ao método *stepwise* para identificar quais variáveis foram mais relevantes na explicação do fenômeno: para cada configuração de ARFV obtida foi aplicado o *stepwise*, baseado no indicador AIC, diferenciando as variáveis que apresentaram nível de significância menor do que 5%. O resultado dessa abordagem foi uma tabela em que cada coluna representa uma configuração de ARFV e cada linha reflete uma variável com a indicação de relevância ou não no modelo/configuração. Para as variáveis mais importantes foram apresentados os coeficientes e desvios padrão da modelagem, além do ajuste global (erro médio absoluto – EMA) e local (gráfico de resíduos por observação). Destaca-se que, por se tratarem de valores categóricos, as variáveis foram inseridas nos modelos como fatores.

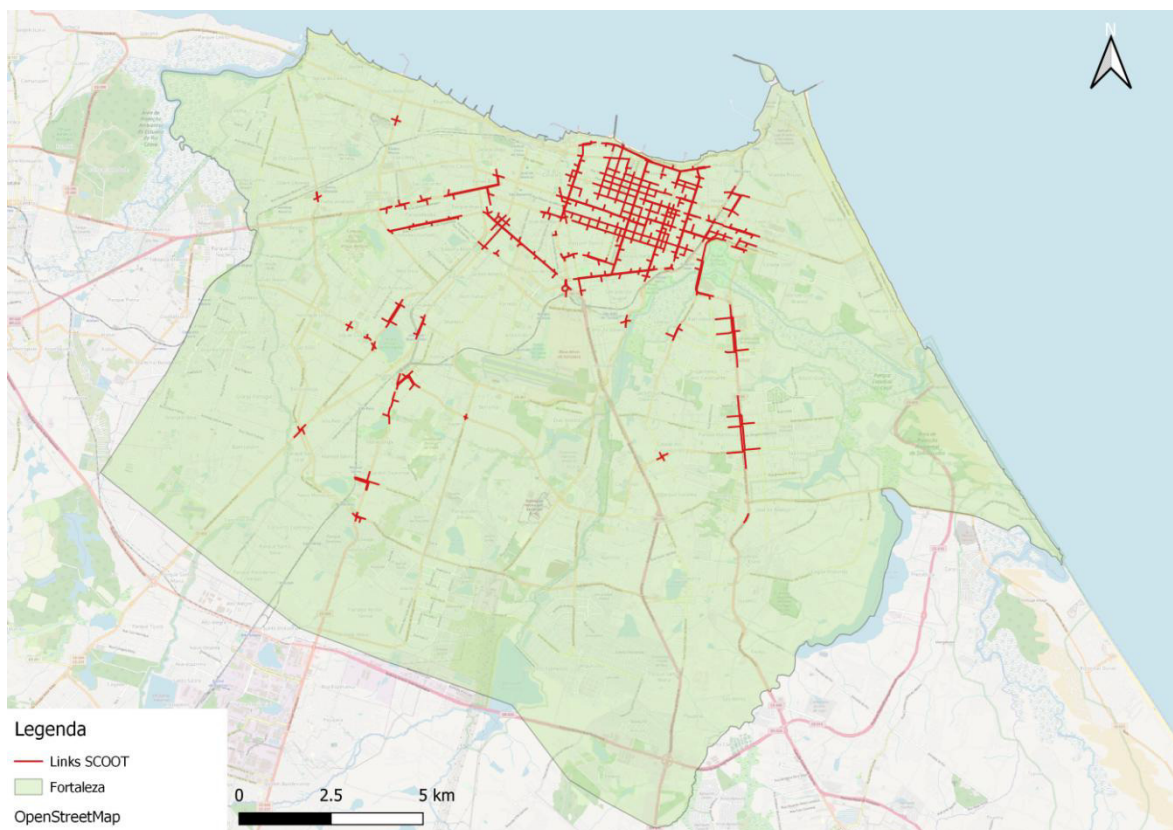
A modelagem por entidades foi realizada utilizando o *banco de dados acidentes + entidade-anual* (3). Novamente as distribuições de Poisson e Binomial Negativa foram utilizadas com a avaliação do  $\sigma_d$ , juntamente com a técnica *forward* de seleção de variáveis. Os indicadores AIC, logaritmo da máxima verossimilhança e EMA foram aplicados na seleção dos modelos mais robustos. Por fim, aplicou-se o *CuRe plot* para análise dos resíduos locais do modelo.

Definidos os modelos estatisticamente mais robustos de cada abordagem, interpretaram-se os parâmetros estimados e realizou-se uma análise gráfica da modelagem por tipo de severidade dos acidentes, buscando o levantamento de hipóteses sobre a relação entre as variáveis consolidadas e os acidentes de trânsito e a possível confirmação da obtenção de resultados mais confiáveis a partir da aplicação de dados desagregados quando comparados a medidas em agregação anual.

## 4 RESULTADOS E DISCUSSÃO

Seguindo as etapas propostas no método, esta seção apresenta os resultados obtidos e a discussão baseada nas hipóteses levantadas na seção da revisão bibliográfica. Como as principais informações de tráfego foram extraídas do sistema *SCOOT*, o objeto de estudo deste trabalho foi as interseções semaforizadas que possuíam o sistema nos anos de 2015 a 2017 (Figura 13). Os dados de acidentes utilizados também foram para o mesmo período.

Figura 13 – Posição dos laços do SCOOT na cidade de Fortaleza



Fonte: Elaborada pelo autor

### 4.1 Estimação do horário mais provável da ocorrência do acidente a partir de perturbações na velocidade da corrente de tráfego

Nesta seção são apresentados e analisados os resultados em três etapas. Primeiramente, uma breve caracterização do banco de dados foi feita. Em seguida, foi realizado o estudo de caso para definição do critério de identificação da perturbação na velocidade mais efetivo. Por fim, foi aplicado o algoritmo com o critério selecionado para todo o banco de dados de acidentes, desenvolvendo uma análise entre as diferenças dos

horários reportados dos acidentes e da ocorrência das oscilações na velocidade e apresentando exemplos de casos em que foi possível a detecção da perturbação na velocidade, além de casos em que não houve sucesso na identificação.

#### ***4.1.1 União do banco de dados de fiscalização eletrônica e acidentes***

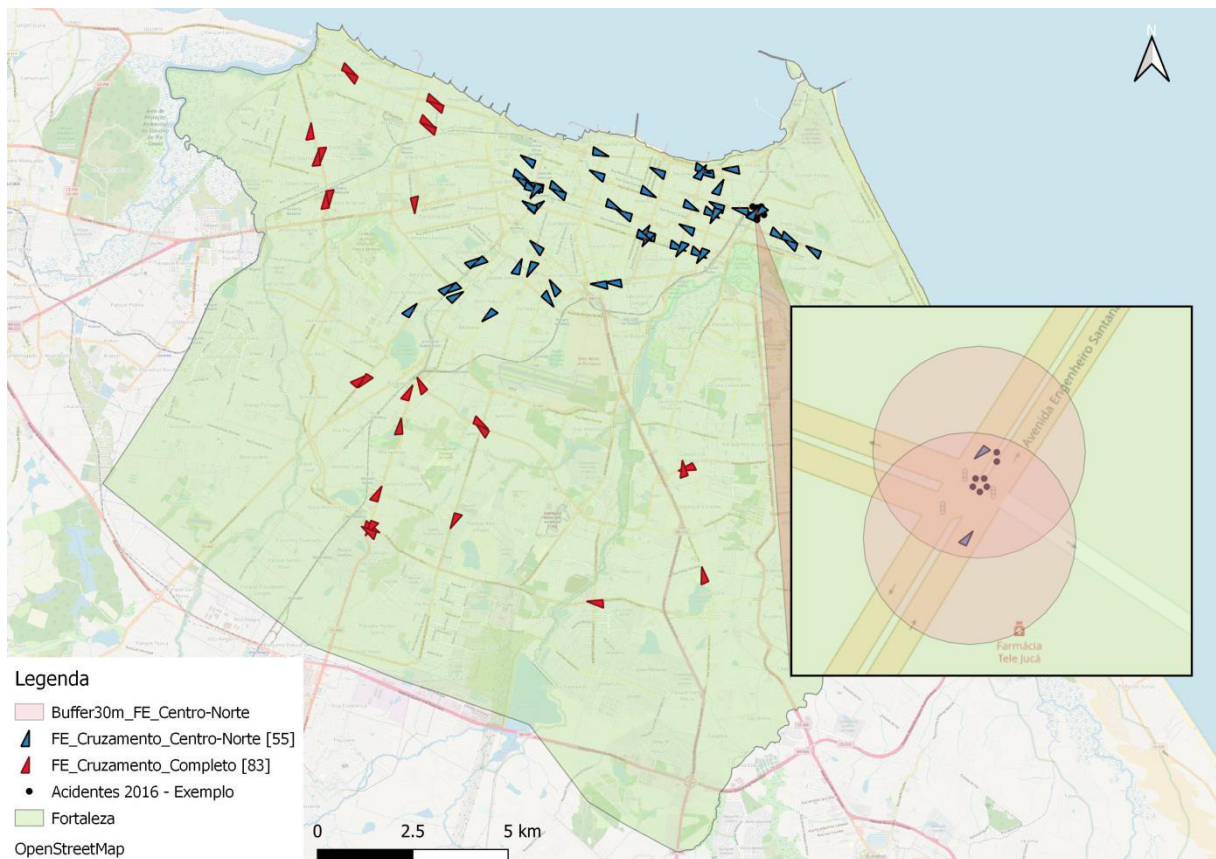
Os equipamentos fixos de fiscalização eletrônica da cidade de Fortaleza são divididos em três principais tipos: interseções semaforizadas, meio de quadra e faixa exclusiva de transporte público coletivo. Como destacado na etapa do método, com o intuito de aumentar a confiança com relação à localização dos acidentes, apenas os equipamentos de interseções semaforizadas foram estudados. Durante o período de 2015 a 2017, 80 equipamentos de fiscalização de cruzamentos estavam ativos em toda a cidade, contemplando 57 interseções. Entretanto, como os laços do *SCOOT* estão presentes basicamente na região centro-norte da cidade (Figura 13), decidiu-se por considerar apenas os cruzamentos nessa localidade, representando 55 equipamentos em 43 interseções semaforizadas.

Dessas 43 interseções, 30 continham equipamento em apenas uma aproximação; 12, em duas aproximações e uma única interseção possuía três dispositivos. A partir de um raio de 30 metros de cada equipamento, pôde-se coletar 329 acidentes em 2015, 369 em 2016 e 215 em 2017; correspondendo a uma amostra de 913 acidentes. A Figura 14 exibe a distribuição espacial dos equipamentos e o destaque na coleta dos acidentes.

#### ***4.1.2 Estudo de caso para definição do critério de identificação de perturbação na velocidade***

A definição do critério de identificação de perturbação na velocidade a ser implementado no algoritmo e da agregação temporal do indicador de velocidade foi realizada com a aplicação de um estudo de caso em 200 acidentes selecionados aleatoriamente do conjunto amostral empregando a combinação de três variáveis em oito diferentes cenários. As variáveis avaliadas foram: (i) agregação temporal da velocidade em 1 e 3 minutos; (ii) média da velocidade no dia típico reduzida de 2 e 3 desvios padrão; (iii) duração da perturbação para detecção de 6 e 9 minutos. Por exemplo, a combinação 1 (Tabela 2) estudou a agregação da velocidade em 1 minuto necessitando que a velocidade permanecesse por, pelo menos, 6 minutos abaixo do valor da média da velocidade reduzida de 2 desvios padrão em dia típico para que a perturbação fosse automaticamente detectada pelo algoritmo.

Figura 14 – Distribuição espacial dos equipamentos de fiscalização eletrônica de interseções semaforizadas e método de coleta dos acidentes



Fonte: Elaborada pelo autor

Para identificação das combinações de melhor desempenho foram consideradas a porcentagem de detecção automática da perturbação na velocidade pelo algoritmo e a porcentagem de perturbações detectadas em que foi possível precisar o horário inicial da perturbação de forma visual. A Tabela 2 traz o resumo dos resultados obtidos.

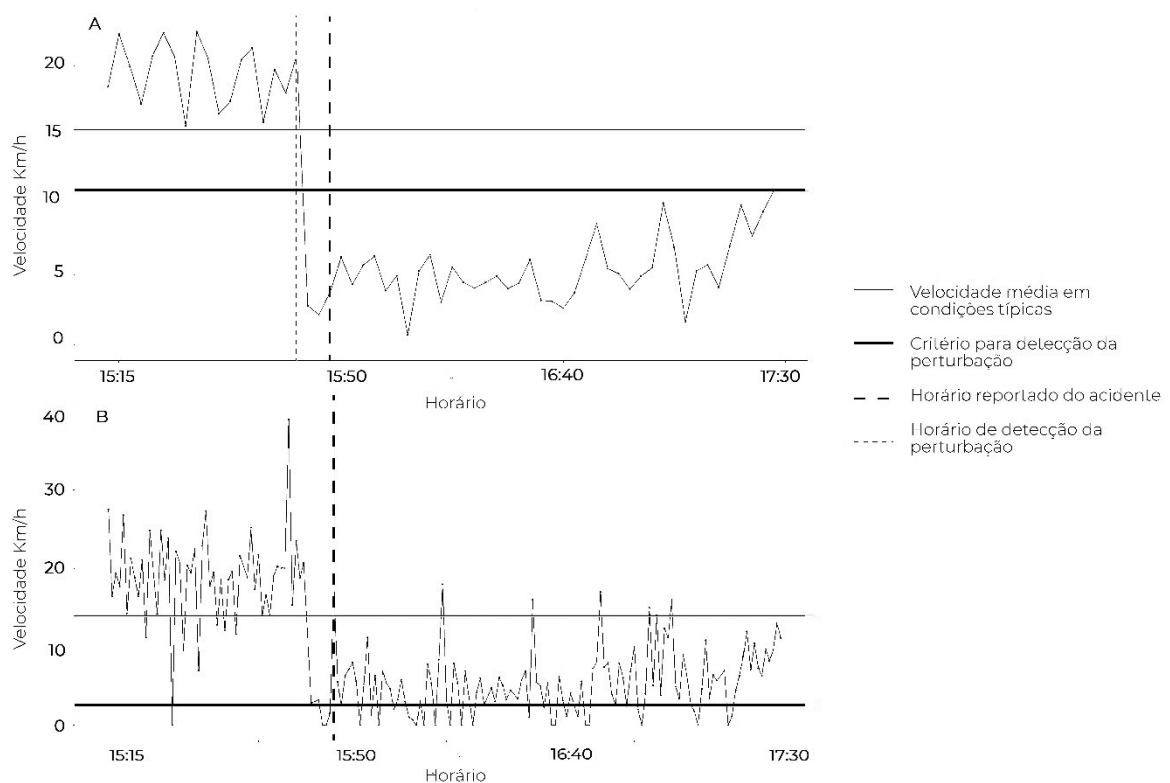
Nota-se que as combinações que empregaram agregação da velocidade de três minutos apresentaram taxas de detecção consideravelmente superiores às agregações de um minuto. A alta oscilação entre velocidades consecutivas agregadas em um minuto – principalmente devido à presença do semáforo – proporcionou altos desvios padrão do indicador, gerando assim, critérios de detecção da perturbação com valores baixos que muitas vezes não foram atingidos mesmo com a significativa redução da velocidade. O agrupamento em três minutos suavizou a oscilação natural causada pelo ciclo semaforístico, reduzindo o desvio padrão entre as observações. A Figura 15 ilustra a situação apresentada acima trazendo a comparação de gráficos de um mesmo acidente com diferença na agregação temporal da velocidade.

Tabela 2 – Resumo do estudo de caso para identificação dos critérios para aplicação do algoritmo de detecção de perturbações na velocidade

ID	Combinação	% Detecção algoritmo	% Confirmação visual
	Tipos		
1	1 min 2 dp 6 dur	10%	70%
2	1 min 3 dp 6 dur	2%	67%
3	1 min 2 dp 9 dur	8%	73%
4	1 min 3 dp 9 dur	1%	100%
5	3 min 2 dp 6 dur	41%	65%
6	3 min 3 dp 6 dur	28%	71%
7	<b>3 min 2 dp 9 dur</b>	<b>36%</b>	<b>73%</b>
8	3 min 3 dp 9 dur	24%	75%

Fonte: Elaborada pelo autor

Figura 15 – Comparação de detecção com durações de três (A) e de um (B) minutos



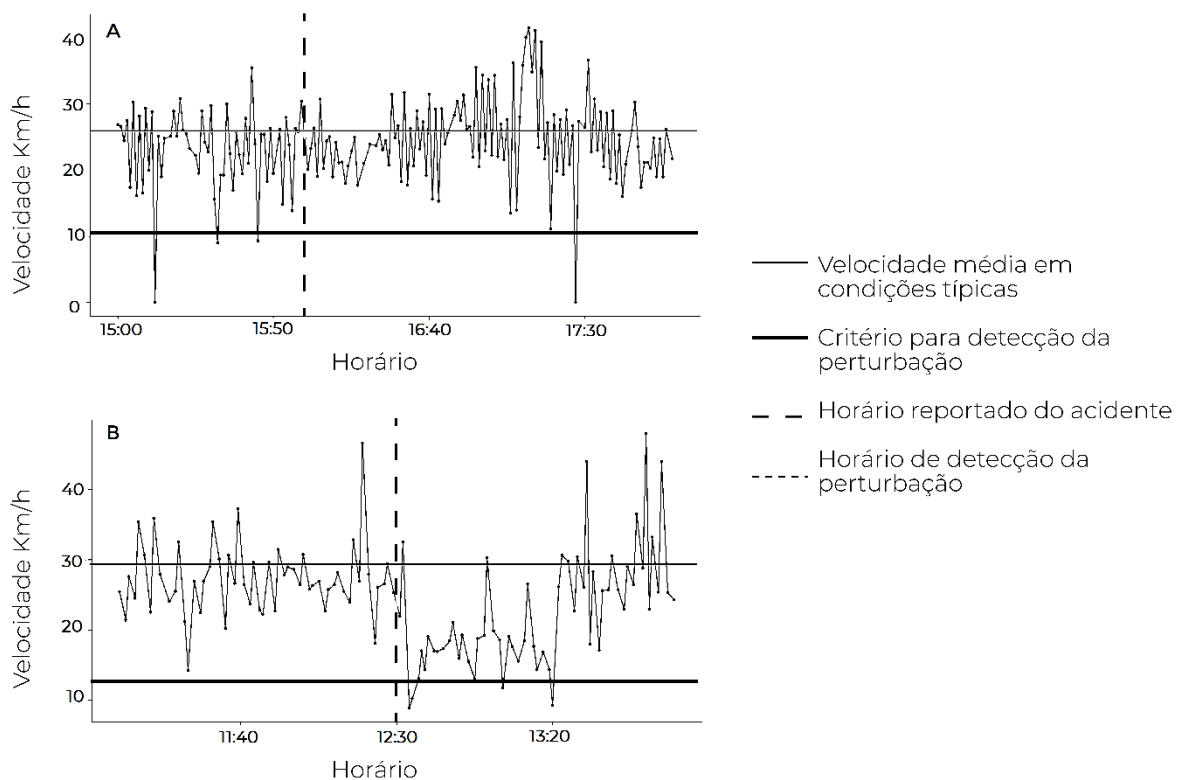
Fonte: Elaborada pelo autor

Além da velocidade não permanecer abaixo do critério de identificação da perturbação, outras hipóteses podem ser levantadas para explicar a não detecção de perturbação em mais da metade dos acidentes: (i) alguns acidentes – embora tenham sido georreferenciados no cruzamento – somente impactaram a aproximação sem fiscalização ou

foram erroneamente georreferenciados pelos agentes de trânsito; (ii) parte dos acidentes ocorreu a montante do laço, assim apenas o indicador de volume seria impactado; (iii) alguns acidentes com menor severidade não impactaram a velocidade na via a ponto de ser detectada pelo algoritmo.

As Figuras 16A (colisão traseira) e 16B (colisão transversal) trazem exemplos em que as perturbações na velocidade não foram identificadas pelo algoritmo. Na colisão traseira percebe-se que não há perturbações significativas no intervalo observado. Já na colisão transversal observa-se uma redução na velocidade após o horário reportado do acidente, mas essa não foi longa nem intensa o suficiente para detecção automática.

Figura 16 – Exemplos de perturbações não detectadas pelo algoritmo: (A) colisão traseira; (B) colisão transversal

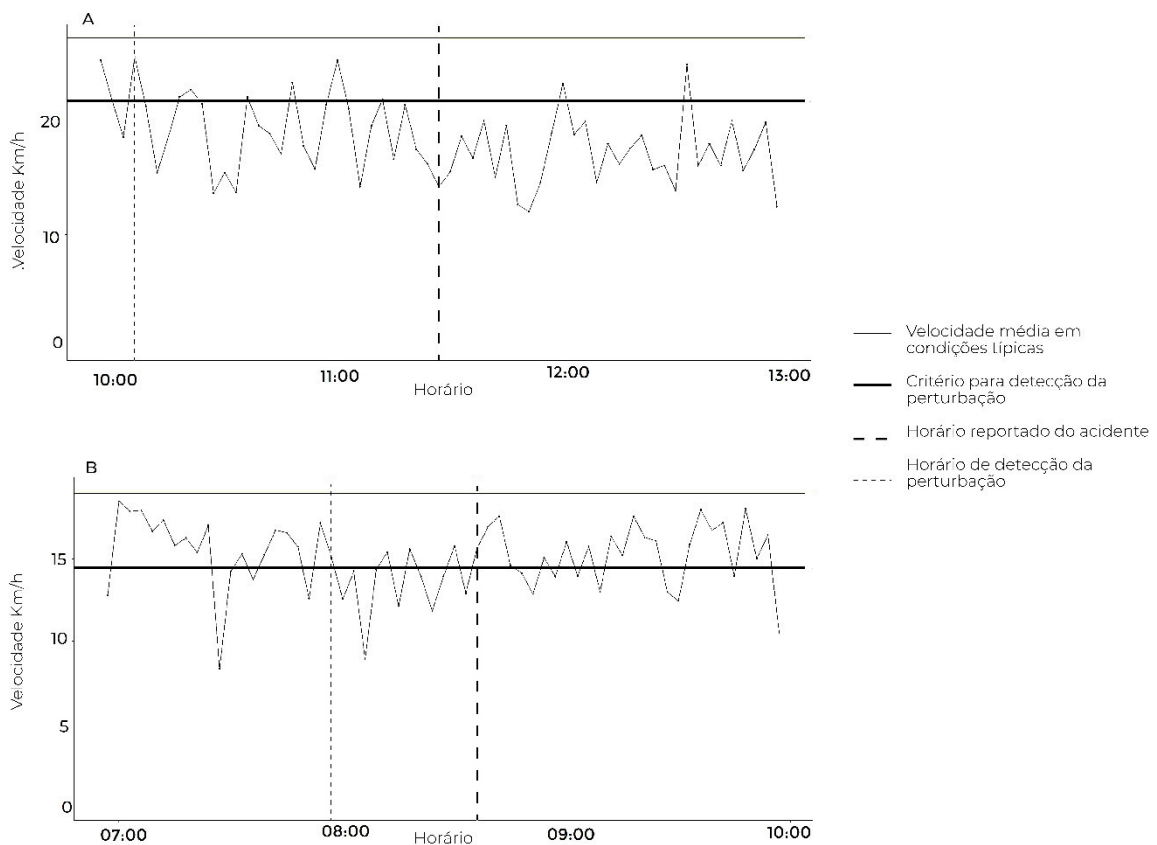


Fonte: Elaborada pelo autor

A Figura 17 traz dois exemplos em que a perturbação na velocidade foi detectada pelo algoritmo, mas – durante a análise visual complementar – não foi possível a determinação do horário inicial da oscilação, dificultando a estimativa do horário de ocorrência do acidente. Na Figura 17A (colisão traseira) o algoritmo detectou uma perturbação por volta das 10h10m,

entretanto como a velocidade não apresentou um padrão de constante de redução, não foi possível precisar o horário de ocorrência do acidente. O perfil de velocidade na Figura 17B (abalroamento lateral) não permitiu associar alguma perturbação pontual à ocorrência do acidente. Uma hipótese para tal comportamento duradouro abaixo da média é de que a redução na velocidade pode ter ocasionado o acidente (uma faixa obstruída por veículos, por exemplo), e não o contrário.

Figura 17 – Exemplos de perturbações detectadas pelo algoritmo, mas que não foram possíveis precisar os horários iniciais das oscilações: (A) colisão traseira; (B) colisão lateral



Fonte: Elaborada pelo autor

Em termos gerais, como esperado, o aumento da duração para nove minutos e a utilização de três desvios padrão reduziu a porcentagem de detecção automática, mas incrementou os casos em que foi possível confirmar visualmente o horário provável dos acidentes. Dentre todas as combinações, a 5 e a 7 foram as que apresentaram melhores resultados. Decidiu-se por adotar a combinação 7 (3 min 2 dp 9 dur) para condução da análise completa devido às maiores taxas de confirmação visual.

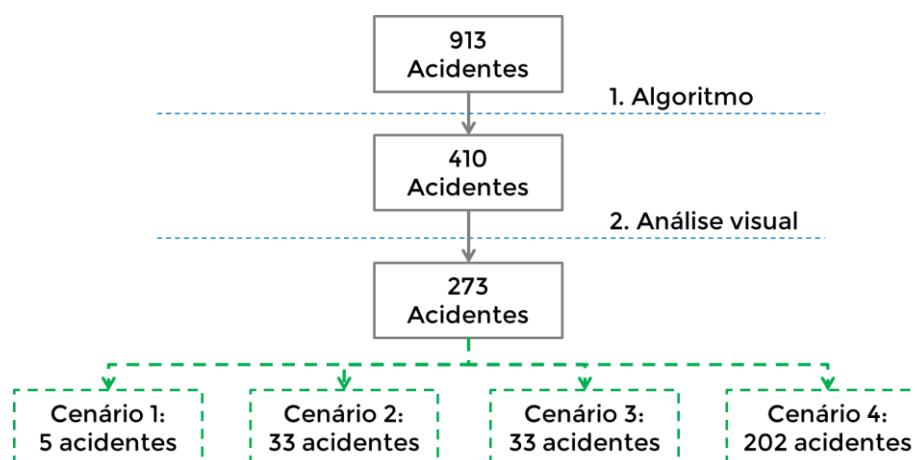


### 4.1.3 Análise das perturbações na velocidade e comparação com os horários reportados dos acidentes

A análise das perturbações de tráfego foi dividida em quatro cenários em função do tipo de acidente e do número de laços que identificaram a perturbação: (i) perturbações detectadas por pelo menos dois laços associadas a colisões transversais; (ii) perturbações encontradas por pelo menos dois laços associadas aos demais tipos de acidentes; (iii) perturbações constatadas por apenas um laço associadas a colisões transversais; (iv) perturbações observadas por apenas um laço associadas aos demais tipos de acidentes. Esta divisão reflete - em ordem decrescente - o nível de confiança para estimativa do horário mais provável do acidente. Por exemplo, o cenário 1 permite que o horário mais provável da ocorrência do acidente seja validado a partir da informação de um dos laços. No cenário 4, por outro lado, a confiança com relação ao georreferenciamento dos acidentes é menor, podendo impactar na qualidade da detecção. Destaca-se que aproximadamente 35% dos acidentes cadastrados no banco de dados não possuem tipificação determinada, sendo assim incluídos na categoria *demais tipos*.

Em termos gerais, dos 913 acidentes investigados, em 410 (44,9%) foram automaticamente identificadas perturbações na velocidade seguindo o critério proposto. Dos acidentes detectados pelo algoritmo, em 273 (66,6%) a análise visual complementar permitiu a detecção exata do momento da perturbação na velocidade, mantendo desempenho próximo ao obtido no estudo de caso. A Figura 18 traz um resumo das detecções por cenário.

Figura 18 – Resumo das detecções das perturbações por etapa e por cenário



Fonte: Elaborada pelo autor

As Figuras 19A e 19B trazem o perfil de velocidade para dois laços detectores no período em que uma colisão transversal foi reportada. Verifica-se que os dois laços detectaram a perturbação na velocidade no mesmo horário (três minutos depois do reportado). As Figuras 19C e 19D referem-se ao intervalo em que uma colisão traseira foi reportada. Observa-se uma diferença de aproximadamente três minutos entre as detecções de cada um dos laços, ocasionada - possivelmente - pela ocorrência do acidente na aproximação da Figura 19C, atrasando o impacto na outra aproximação. Grande parte das perturbações detectadas e validadas visualmente para os cenários 1 e 2 seguiram o perfil apresentado na Figura 19, possibilitando a precisa identificação do horário da ocorrência.

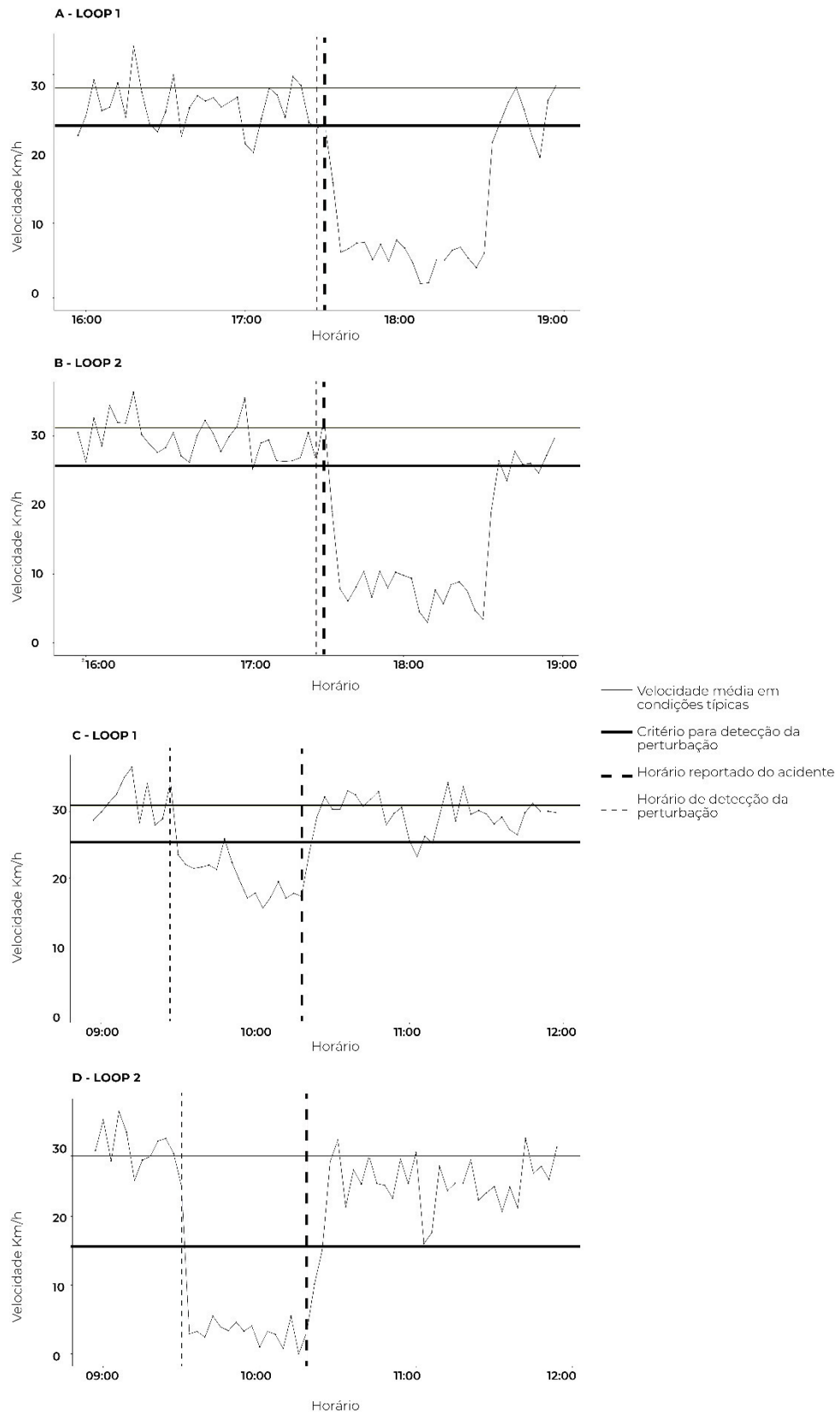
A Figura 20 apresenta duas perturbações detectadas e confirmadas pela análise visual em apenas um laço. As Figuras 20A (colisão transversal) e 20B (colisão lateral) mostram situações em que foi possível determinar precisamente o momento de ocorrência da perturbação na velocidade.

A Tabela 3 traz as diferenças observadas entre o horário reportado do acidente e a perturbação na velocidade por cenário. Para o caso de maior confiança entre a diferença (cenário 1), verificou-se um atraso médio de 27 minutos em relação ao horário reportado no banco de dados e a ocorrência da oscilação na velocidade, entretanto a amostra para esse cenário foi de apenas cinco acidentes. Considerando todos os 273 acidentes, a diferença foi menor: atraso médio de 19 minutos.

Visando refinar os resultados obtidos por característica do acidente, avaliou-se as diferenças por (i) severidade do acidente e (ii) faixa de horário de ocorrência do acidente. Para a severidade do acidente levantou-se a hipótese de que acidentes com vítimas são tratados com maior atenção por parte dos agentes de trânsito do que os com apenas danos materiais, sugerindo que as diferenças entre os horários reportados e a detecção da perturbação foram menores para os acidentes mais severos. A realização de um teste de hipótese unicaudal rejeitou a hipótese nula de que as amostras vêm de uma população de mesma média (valor-p = 0,01), confirmando a hipótese levantada.

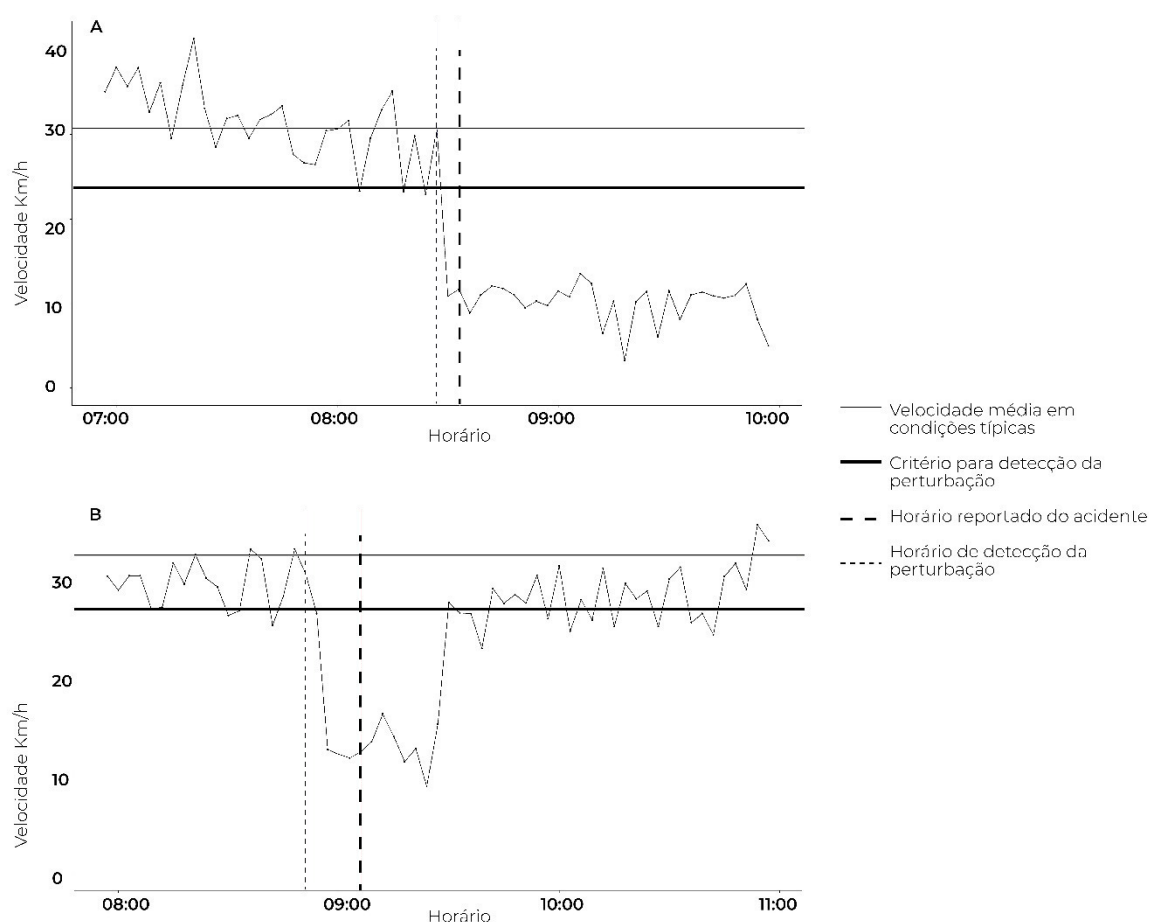
Com relação à faixa horária de ocorrência dos acidentes, foram segregados dois intervalos horários: comercial (8h00m-17h59m) e não comercial (18h00m-7h59m). O levantamento de hipóteses para este caso foi de maior complexidade, visto que tanto o contingente de agentes de trânsito, que é maior no intervalo comercial, quanto a demanda veicular, que é menor no intervalo não comercial, podem impactar nas diferenças estimadas. O teste de hipótese bicaudal não rejeitou a hipótese nula de que as amostras vêm de uma população de mesma média (valor-p = 0,17).

Figura 19 – Perturbações detectadas por dois laços em horários similares: (A) e (B) colisão transversal; (C) e (D) colisão traseira



Fonte: Elaborada pelo autor

Figura 20 – Perturbações detectadas por apenas um laço: (A) colisão transversal; (B) colisão lateral



Fonte: Elaborada pelo autor

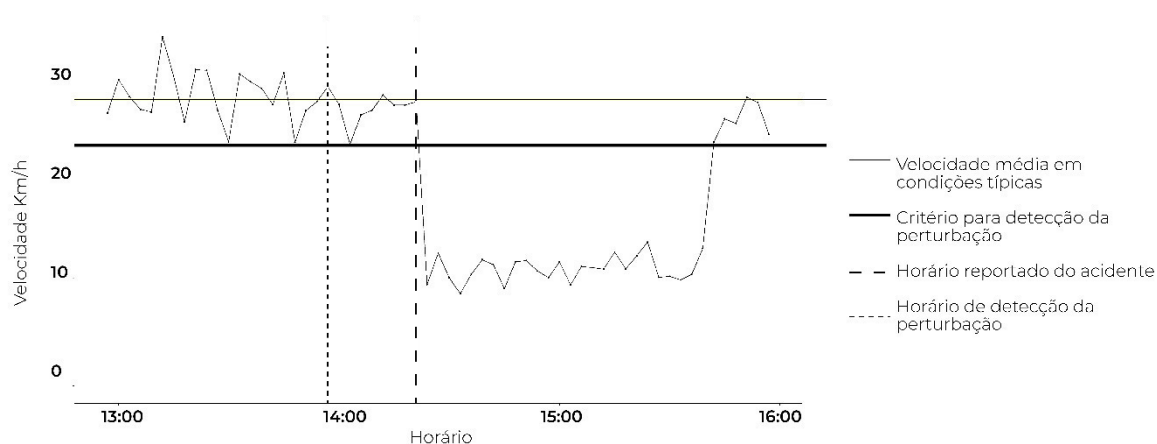
Tabela 3 – Estatísticas entre as diferenças entre o horário reportado dos acidentes e a perturbação na velocidade por cenário

Tipo	N	Estatísticas das diferenças em minutos							
		Média	Desv. Pad.	Mín	25° Perc	75° Perc	95° Perc	Máx	
Por cenário	1	5	27	36	-6	3	42	76	84
	2	33	26	25	-12	9	48	70	75
	3	33	11	19	-30	0	18	47	60
	4	202	18	20	-27	4	30	63	87
Por severidade	Danos materiais	179	21	23	-30	6	33	66	87
	Vítimas	94	15	18	-15	3	24	44	84
Por horário	8h00m-17h59m	191	20	23	-30	6	33	67	87
	18h00m-7h59m	82	16	19	-12	3	21	57	84
Geral	273	19	22	-30	3	30	63	87	

Fonte: Elaborada pelo autor

Ao avaliar os percentis iniciais das diferenças estimadas, constata-se que em todos os casos houve acidentes reportados antes da identificação da perturbação, apresentando diferença estimada menor que zero, sugerindo a ocorrência de falhas no cadastro dos horários dos acidentes no banco de dados do SIATFOR. A Figura 21 traz um exemplo desse caso.

Figura 21 – Exemplo de acidente reportado antes da detecção da perturbação na velocidade (colisão transversal)



Fonte: Elaborada pelo autor

Os resultados obtidos nesta seção foram utilizados na “correção” dos horários reportados dos acidentes durante a etapa de consolidação de banco dos dados. As correções que foram aplicadas nos horários dos acidentes foram baseadas no valor do 95º percentil das diferenças observadas, assim, assumindo que a amostra de acidentes estudada reflete às condições observadas em toda a população; garantiu-se que em aproximadamente 95% das situações os regimes de fluxo veicular foram atribuídos a condições precursoras aos acidentes. Como foi verificada diferença estatisticamente significativa a um nível de confiança de 1% entre as médias amostrais das diferenças por severidade do acidente, decidiu-se por segmentar as “correções” por severidade: acidentes com apenas danos materiais: 66 minutos; e acidentes com vítimas: 44 minutos.

Sabe-se da limitação dessa abordagem, visto que para os acidentes em que o horário de reportagem está correto ou adiantado em relação à identificação da perturbação são coletadas informações de tráfego que não representam as condições imediatamente precursoras. A utilização apenas dos acidentes em que puderam ser precisados os horários de ocorrência inviabilizaria a modelagem devido ao baixo número de observações, visto que apenas quatro interseções possuem o *SCOOT* e a fiscalização eletrônica. Resumindo, para

garantir a utilização de condições precursoras em 95% dos acidentes, há de se conviver com características de tráfego não imediatamente precursoras para certas observações. Entretanto, como o principal objetivo deste trabalho é avaliar condições macroscópicas de regimes veiculares, e não oscilações microscópicas no fluxo veicular, entendeu-se que essa limitação não inviabiliza a obtenção de resultados robustos e confiáveis. Na seção 4.3.4, apresenta-se uma análise de sensibilidade para avaliar o impacto da utilização da “correção média” nos resultados.

## **4.2 Consolidação dos bancos de dados**

Nesta seção são apresentados e explorados os bancos de dados necessários para desenvolvimento das etapas de definição dos agrupamentos e da modelagem. Inicialmente, foi apresentado o banco de dados de RFV, sendo, em seguida, exibida a associação desse banco com os acidentes de trânsito e, por fim, expôs-se o banco de dados de entidades-anuais.

Recordando, os regimes de fluxo veicular (RFV) são compostos por atributos fixos, que consistem na geometria e no tipo de operação da interseção semaforizada, e por atributos variáveis, representados pelas características de tráfego que possuem oscilação dentro do período de um mesmo dia. Uma vez que os regimes são determinados, abandona-se a idealização de características pertencentes à entidade  $i$ , partindo para uma concepção desagregada baseada na particularidade daquele regime.

### ***4.2.1 Consolidação do banco de dados de regimes de fluxo veicular***

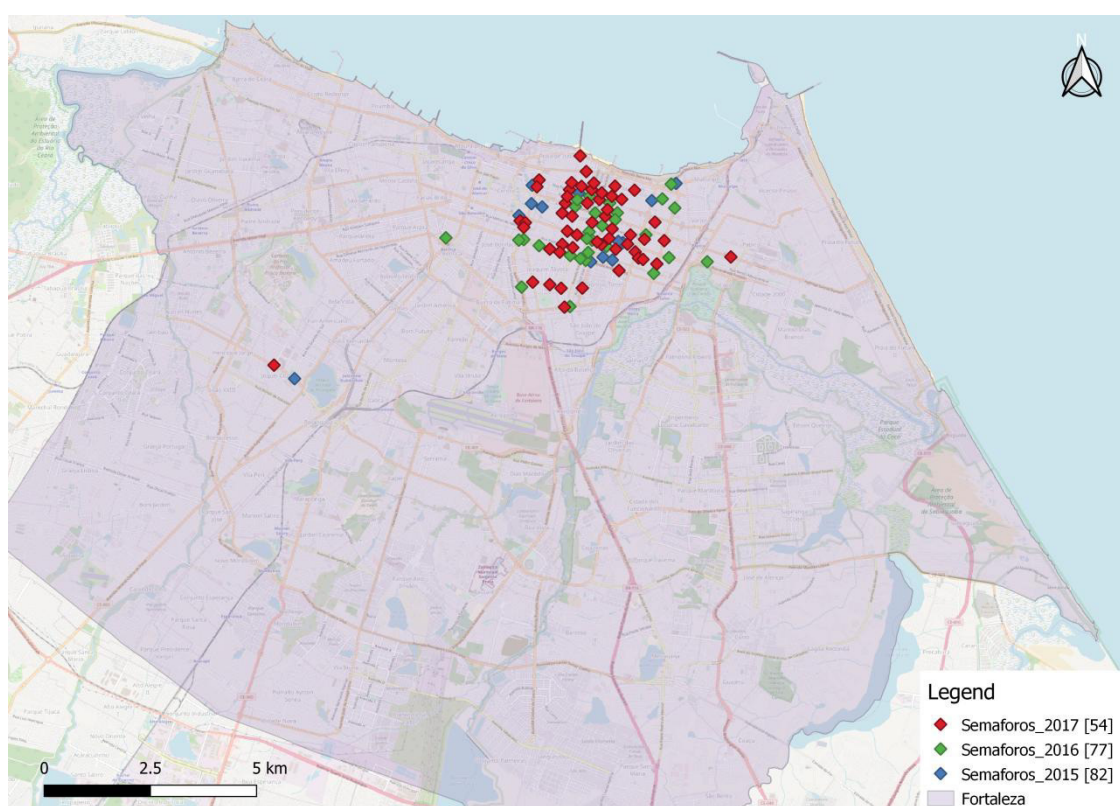
As variáveis de tráfego foram coletadas do registro de observações do sistema *SCOOT* em agregações de 15 minutos. Visando eliminar locais que apresentaram frequentes falhas na operação dos laços, adotou-se que para uma interseção ser considerada ela deveria conter pelo menos 50% de dados de tráfego válidos em todas as aproximações durante o período completo avaliado. Por exemplo, para o ano de 2015 o período completo contempla 35.040 intervalos de 15 minutos (365 dias \* 24 horas \* 4 intervalos). Assim, foram consideradas 82, 77 e 54 interseções semaforizadas nos anos de 2015, 2016 e 2017; respectivamente. A Figura 22 exhibe a localização dos semáforos considerados no estudo.

Percebe-se que a grande maioria dos semáforos que possuem o sistema *SCOOT* está na região centro-norte da cidade. Essa região é caracterizada pelo adensamento de comércios,

de serviços e de residências verticais, gerando grande movimentação de veículos, pedestres e ciclistas durante todo o dia; além de ser a zona de cidade de maior adensamento semafórico.

Das diversas informações coletadas pelo *SCOOT*, decidiu-se trabalhar com o *fluxo veicular por faixa* – que representa a estimativa do fluxo horário veicular para cada 15 minutos –, a *velocidade*, o *grau de saturação* e o *número de paradas veiculares*. A partir do fluxo veicular pôde-se obter a relação entre os fluxos nas direções principais e secundárias, hierarquizadas de acordo com o fluxo.

Figura 22 – Interseções semaforizadas contempladas no estudo



Fonte: Elaborada pelo autor

Uma das principais atribuições do *SCOOT* é determinar, a partir da demanda veicular, o tempo de verde para cada uma das aproximações semafóricas, entretanto essa informação estava presente apenas em cerca de 20% das observações registradas pelo sistema, inviabilizando sua obtenção via *SCOOT*. Assim, para estimar a *duração do ciclo semafórico* utilizou-se a programação diária fixa empregada pelos gestores da operação semafórica da cidade, que é baseada em dados históricos do *SCOOT*.

Após aplicar todos os filtros e unir os dados para os três anos do estudo, obteve-se uma amostra de 4.102.567 intervalos de 15 minutos de informações de tráfego, valor que compreende aproximadamente 55% do intervalo temporal completo.

Os atributos fixos (*número de faixas, número de aproximações semaforicas e número de movimentos de conversão permitidos*) foram coletados a partir de bases georreferenciadas fornecidas pela Prefeitura e de consultas ao *Google Earth*. A partir de um raio de 100 metros da interseção foi coletado o número de equipamentos de fiscalização eletrônica. Esses atributos foram associados aos intervalos com as informações de tráfego, finalizando a consolidação do *banco de dados RFV – 15 min*.

Em uma primeira caracterização do banco de dados de RFV, apresentam-se as estatísticas de tendência central, dispersão e forma dos dados coletados com foco nos atributos variáveis (Tabela 4). Destaca-se que nas próximas seções também são realizadas outras análises exploratórias, de acordo com a necessidade emergente.

Tabela 4 – Estatística descritiva dos regimes de fluxo veicular

Variável	Média	Desvio padrão	Mín	25° Perc	75° Perc	Máx	%: = 0*	%: >= 1**
Fluxo por faixa (vei/h/fx)	297	144	10	178	418	768	-	-
Velocidade (km/h)	19	6	1	13	20	55	-	-
Grau de saturação (%)	47	26	1	28	62	250	-	-
Número de paradas por faixa (par/h/fx)	189	111	1	97	273	728	-	-
Duração do ciclo semaforico (seg)	98	30	40	72	128	160	-	-
Relação fluxo via principal-secundária	2,1	1,6	1,0	1,2	1,6	20,0	-	-
Fiscalização eletrônica (quantidade em raio de 100m)	0,05	0,24	0	0	0	2	95%	5%
Número de faixas	5,3	1,4	4	4	6	10	-	-
Número de aproximações	2,5	0,6	2	2	3	5	-	-
Número de conversões permitidas	2,47	0,7	1	2	3	6	-	-

\*: % de RFV sem presença da variável; \*\*: % de RFV com presença da variável

Fonte: Elaborada pelo autor

Em termos gerais, a Tabela 4 mostra que o *fluxo por faixa* abrangeu regimes com diversas demandas veiculares, oscilando, geralmente, entre 150 e 450 veículos por hora por



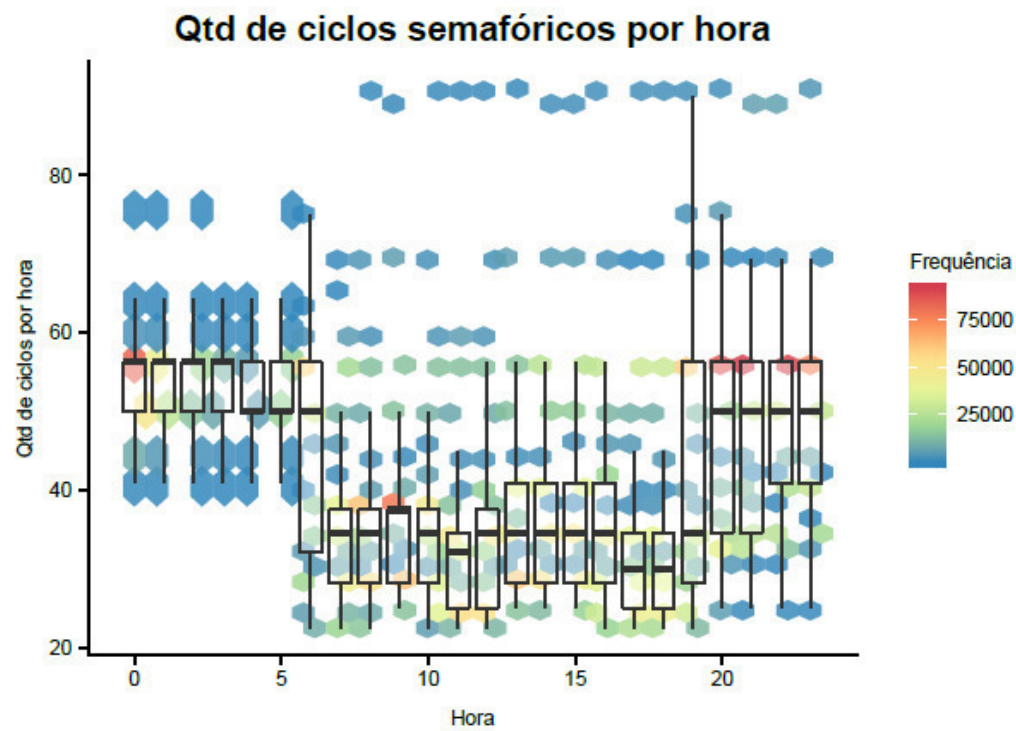
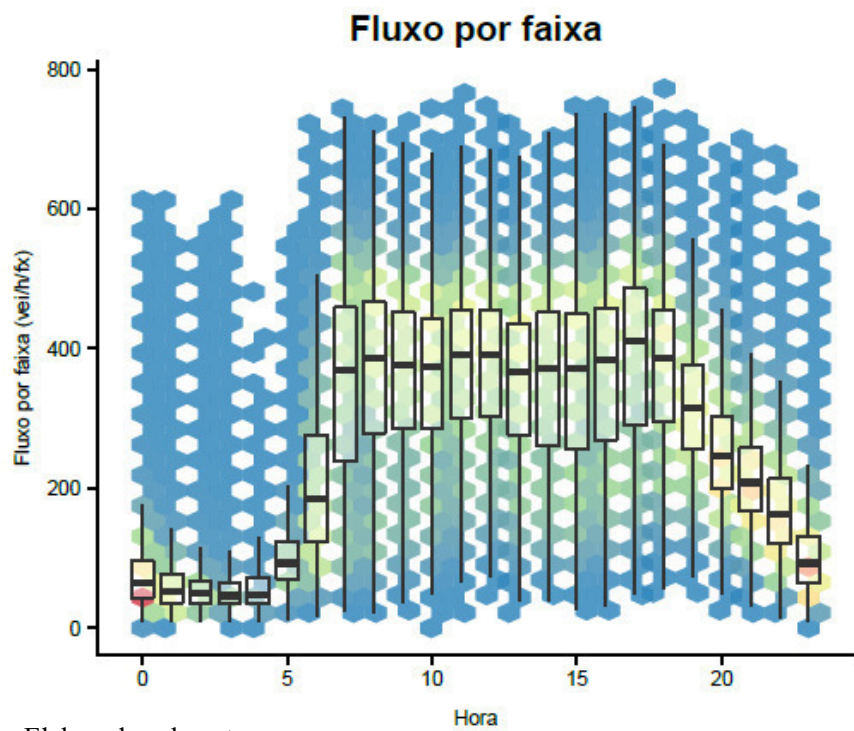
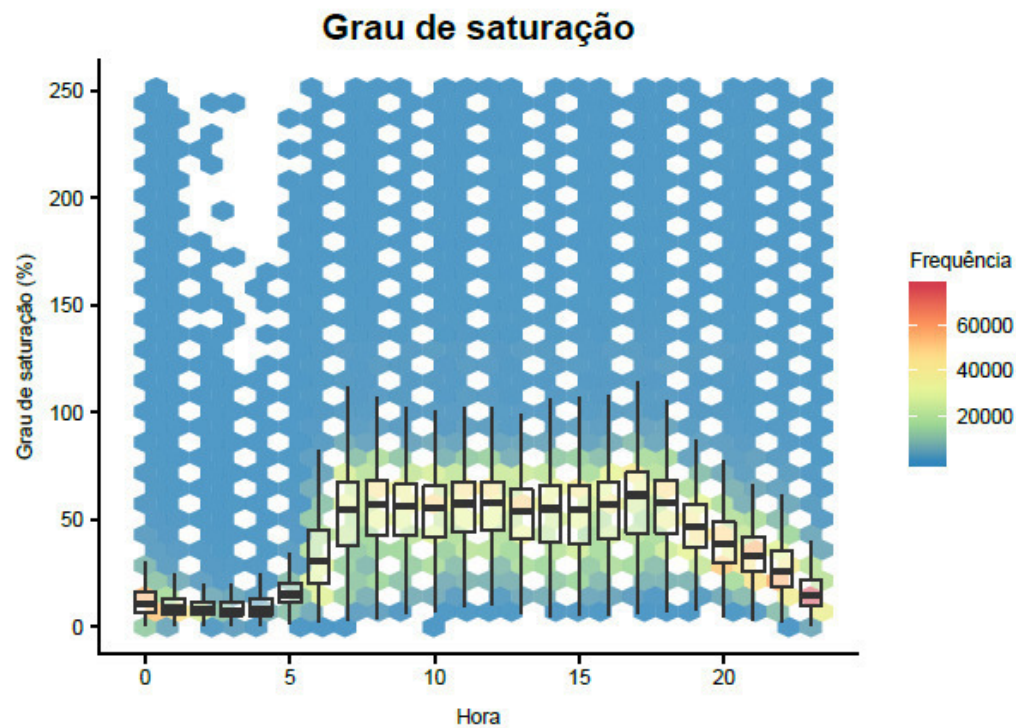
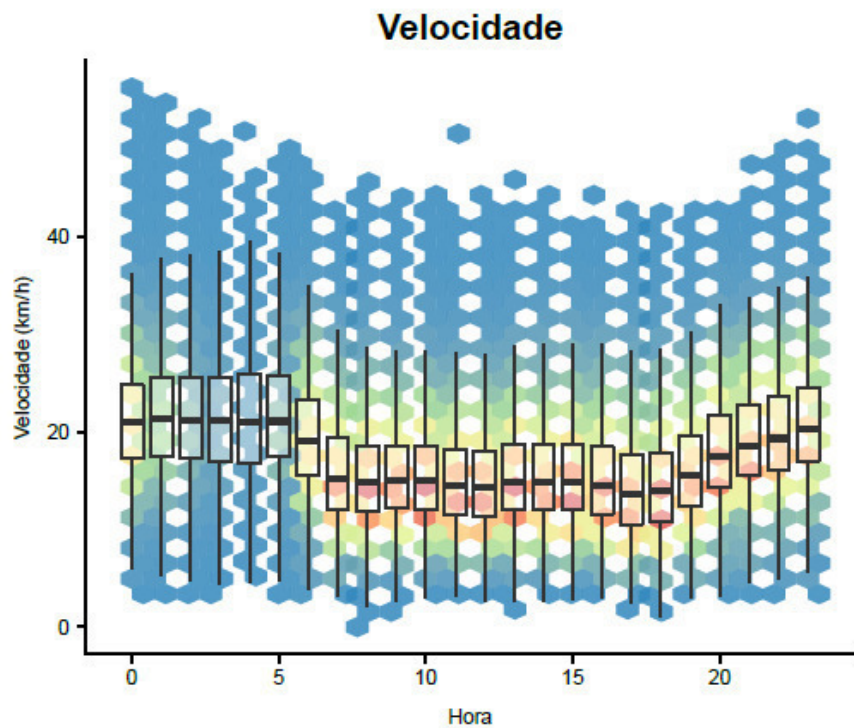
faixa (vphpf). Como esperado, devido aos altos volumes veiculares na região de presença dos laços, as *velocidades* médias estimadas foram baixas, com valores mínimos de 1 km/h, representando regimes altamente congestionados (relembra-se que a *velocidade* não é uma das variáveis diretamente medida pelo *SCOOT* e sim estimada a partir do atraso veicular médio e do tempo de viagem em fluxo livre na aproximação). O *grau de saturação*, que representa a relação entre a demanda veicular, o fluxo de saturação e o tempo de verde efetivo da aproximação semaforica; normalmente flutuou entre 40% e 80%, apresentando também casos acima da capacidade da aproximação. Os *ciclos* estabelecidos na programação semaforica variaram de 48s a 160s, e a relação entre o *fluxo principal-secundária* indicou que, em média, a via principal concentrou quase o dobro do fluxo veicular em relação à secundária.

A estatística dos atributos fixos mostra que em apenas 5% dos RFV havia a presença de equipamentos de *fiscalização eletrônica*, podendo dificultar a avaliação da relação dessa variável com os acidentes. Verifica-se uma concentração de interseções com 4 a 6 *faixas de tráfego* e 2 a 3 *aproximações e conversões permitidas*. A Figura 23 traz imagens com boxplots e frequências da distribuição das principais variáveis de tráfego durante as 24 horas do dia. Substituiu-se o *ciclo semaforico* pela variável *quantidade de ciclos por hora*, visando facilitar a interpretação da variável junto à frequência de acidentes e às características de tráfego.

Verifica-se na Figura 23 que as variáveis *fluxo por hora*, *grau de saturação* e *ciclos por hora* possuíram sensíveis oscilações durante as horas do dia, enquanto a *velocidade* variou em menores escalas. Como esperado, em horários não comerciais a demanda veicular foi menor, ocasionando baixos *fluxo por faixa* e *grau de saturação* e permitindo o desenvolvimento de maiores velocidades. Nota-se, também, que a programação semaforica foi consideravelmente correlacionada à demanda veicular, em que ciclos mais longos ocorreram, principalmente, quando o *fluxo por faixa* foi maior.

#### **4.2.2 Consolidação do banco de dados de acidentes associados aos regimes de fluxo veicular**

Os acidentes de trânsito foram coletados a partir de um raio de 50 metros de cada interseção semaforizada. Para cada acidente foram obtidas as seguintes informações: (i) severidade: vítima ou apenas danos materiais; (ii) tipo: colisão traseira, transversal, lateral, frontal, entre outros; (iii) características da ocorrência para associação aos RFV: data, horário reportado e código da interseção. Nesta etapa foram aplicados os resultados obtidos no



objetivo I, em que os horários reportados foram “corrigidos” a partir dos 95º percentis encontrados para os acidentes com vítimas e danos materiais. Como só foi possível obter informações de tráfego para 55% dos RFV que compõe o espaço temporal completo, não foi possível associar alguns acidentes a RFV, os quais foram descartados da análise. A Tabela 5 traz um resumo do processo de obtenção dos acidentes e da associação aos RFV.

Tabela 5 – Resumo do processo de obtenção dos acidentes e da associação aos RFV

Ano	#Interseções	Acidentes raio 50m	Acidentes associados a RFV
2015	82	412	289
2016	77	379	247
2017	54	203	132
Geral	-	994	668

Fonte: Elaborada pelo autor

Dos 994 acidentes coletados de cada interseção, 668 (67%) puderam ser associados a condições de RFV, compondo o número de observações a ser utilizado nas análises exploratórias e na modelagem. A Tabela 6 traz um resumo dos tipos e severidade dos 668 acidentes.

Tabela 6 – Resumo dos acidentes associados a RFV por tipo e severidade

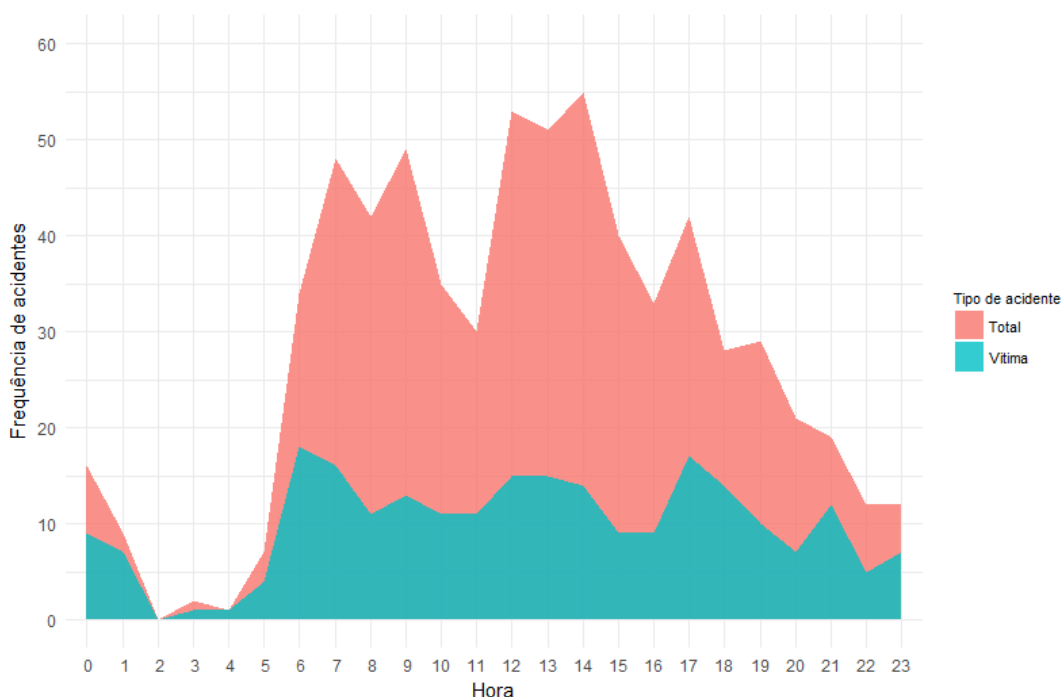
Tipo de acidente	Severidade	
	Vítima	Danos Materiais
Atropelamento	25	3
Capotamento	3	1
Choque c/ obstáculo fixo	5	11
Colisão frontal	4	6
Colisão lateral	32	100
<b>Colisão transversal</b>	<b>47</b>	<b>80</b>
<b>Colisão traseira</b>	<b>18</b>	<b>93</b>
Engavetamento	4	7
Queda	9	4
Tipo não identificado	89	127
<b>Soma</b>	<b>236</b>	<b>432</b>

Fonte: Elaborada pelo autor

Com relação à severidade dos acidentes, 35% apresentaram vítimas e 65%, apenas danos materiais. A respeito do tipo, percebe-se a que as colisões laterais, transversais e traseiras foram as de maior frequência. O elevado número de acidentes com tipo não identificado ocorreu, principalmente, devido aos atendimentos dessas ocorrências serem realizados por órgãos que não coletam esse tipo de informação (SAMU, por exemplo).

Destaca-se, ainda, que dos 236 acidentes com vítimas, 153 envolveram pelo menos um motociclista e 25 foram do tipo atropelamento, indicando que a ocorrência dos acidentes com vítimas pode estar mais relacionada ao tipo de usuário do meio de transportes do que a condições de RFV. Essa hipótese será melhor avaliada nas próximas seções. A Figura 24 apresenta a distribuição horária dos acidentes totais e com vítimas.

Figura 24 – Distribuição horária dos acidentes totais e com vítimas



Fonte: Elaborada pelo autor

A distribuição horária dos acidentes mostra que a partir das 20h00m às 5h59m aproximadamente 50% dos acidentes resultaram em vítimas, intervalo horário caracterizado por menores fluxos veiculares e maiores velocidades, como apresentado na Figura 23. Já para o período das 6h00m às 19h59m, que representa RFV mais demandados, a proporção de acidentes com vítimas foi reduzida para cerca de 25%. Corroborando com a hipótese 1 de que o incremento do fluxo veicular nos intervalos de 15 minutos está associado a maiores frequências de acidentes e à redução da severidade das ocorrências. A Tabela 7 traz a estatística descritiva dos RFV associados a acidentes, além de apresentar as estatísticas para todos os 4,1M de RFV visando comparar a distribuição das duas considerando o efeito da exposição.

Em termos gerais, verifica-se que a média e percentis do *fluxo por faixa*, do *grau de saturação* e do *número de paradas por faixa* foram superiores para os RFV que ocorreram

acidentes em comparação a todos os RFV. Consequentemente observam-se indicadores de *velocidade* suavemente inferiores nos RFV com acidentes. As outras variáveis não exibiram diferenças significativas entre os intervalos com e sem acidentes. As estatísticas apresentadas para os RFV com frequência de acidentes contemplam todos os acidentes – sem distinção por severidade ou tipo –, espera-se que com a segmentação dos acidentes sejam notadas peculiaridades significativas nos RFV por categoria de ocorrência.

Tabela 7 – Estatística descritiva dos RFV associados a acidentes

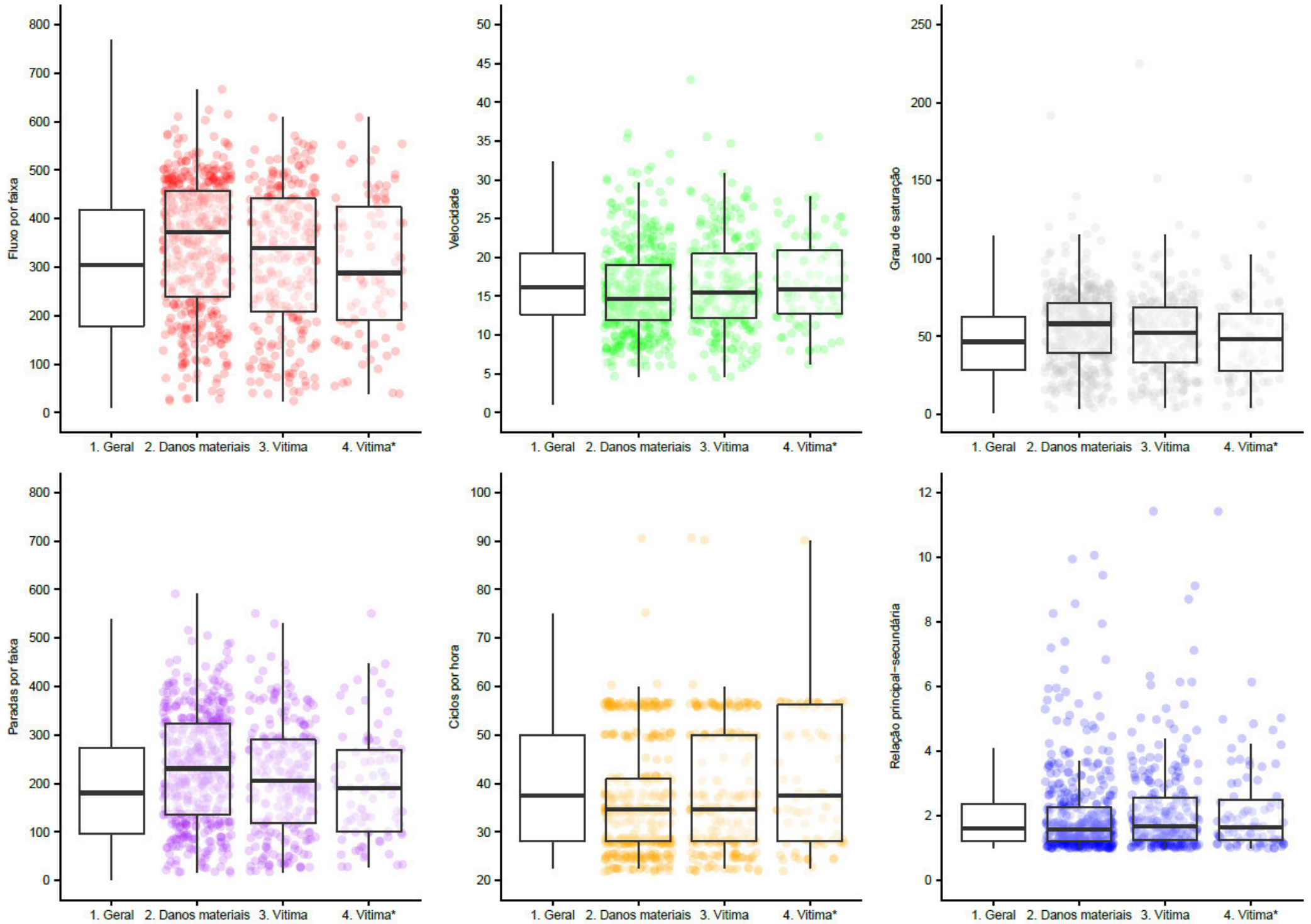
Variável	RFV - Geral				RFV - Acidentes			
	Média	Desvio padrão	25° Perc	75° Perc	Média	Desvio padrão	25° Perc	75° Perc
Fluxo por faixa	297	144	178	418	337	140	225	452
Velocidade	19	6	13	20	16	6	12	19
Grau de saturação	47	26	28	62	54	25	37	70
Número de paradas por faixa	189	111	97	273	224	115	127	315
Ciclos semafóricos por hora	40	12	28	50	37	12	28	50
Relação fluxo via principal-secundária	2,1	1,6	1,2	1,6	2,1	1,4	1,2	2,4
Fiscalização eletrônica	0,05	0,24	0	0	0,05	0,25	0	0
Número de faixas	5,3	1,4	4	6	5,8	2,0	4	6
Número de aproximações	2,5	0,6	2	3	2,7	0,7	2	3
Número de conversões permitidas	2,5	0,7	2	3	2,7	0,9	2	3

Fonte: Elaborada pelo autor

A Figura 25 traz os boxplots das variáveis de tráfego segmentando os acidentes por severidade. Como a maior parte dos acidentes com vítimas ocorreu com usuários vulneráveis (64,8% com motociclistas e 10,6% com pedestres), decidiu-se considerar uma categoria de acidentes com vítimas com usuários de veículos de pelo menos quatro rodas (66 acidentes), visto que a vulnerabilidade dos motociclistas e pedestres pode ser fator preponderante para a severidade do acidente.

O gráfico das variáveis *fluxo por faixa* e *paradas por faixa* mostram tendência de redução da gravidade do acidente com o aumento do valor da variável. Verifica-se, também, que os quartis dessas variáveis para os acidentes com danos materiais foram superiores aos do boxplot geral, indicando que os RFV mais carregados foram mais relacionados à ocorrência

Figura 25 - Boxplot RFV associados aos acidentes por severidade



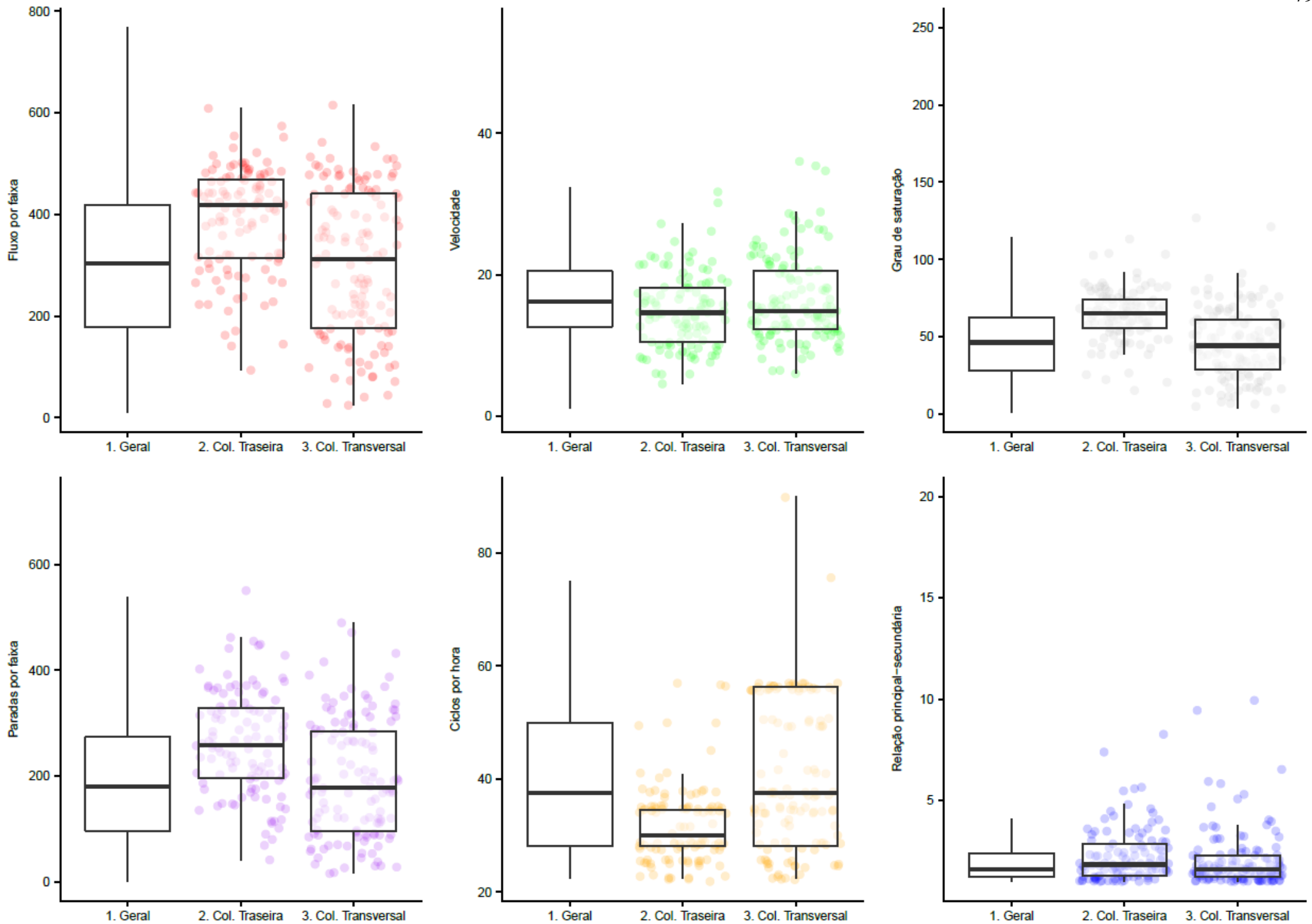
de acidentes de baixa severidade (hipótese 1). O *grau de saturação*, que une a demanda veicular e a capacidade da aproximação, também apresenta comportamento semelhante às citadas acima, corroborando que os RFV mais demandados foram os de maior probabilidade de ocorrência de acidentes, principalmente menos severos. A segregação dos acidentes com vítimas exhibe que as vítimas “não vulneráveis” tenderam a ocorrer em RFV menos carregados, o que era esperado, dado a proteção e os mecanismos de segurança dos veículos de quatro rodas, exigindo maiores velocidades para ocorrência de impactos prejudiciais aos condutores e passageiros.

Apesar de apresentar tendência de aumento da severidade do acidente com o incremento do número de *ciclos por hora*, a existência de alta correlação entre variável e a demanda veicular (Figura 23) dificultou a extração de assertivas baseadas em uma análise univariada. Supõe-se que RFV com altos valores de *ciclo por hora* são mais propícios à ocorrência de acidentes, visto que existe aumento nas interações traseiras – geralmente de baixa severidade – e transversais – usualmente mais severas; mas o fato de os ciclos semaforizados serem mais longos em regimes de maior demanda veicular impossibilita que conclusões robustas sejam efetivadas. Não foi possível verificar tendências no boxplot da variável *relação fluxo principal-secundária*. Destaca-se que nas próximas seções são apresentadas análises bivariadas, permitindo que o efeito do *ciclo por hora*, por exemplo, seja avaliado em conjunto com o *fluxo por faixa*. A Figura 26 traz os boxplots segregados por tipo de acidente.

O estudo das colisões traseiras quase sempre é relacionado à análise dos acidentes com danos materiais, visto que os conflitos longitudinais geralmente não trazem severidade significativas aos usuários. Neste estudo não foi diferente. Das 111 colisões traseiras avaliadas, 93 (83,7%) ocasionaram apenas danos materiais. Os boxplots de *fluxo por faixa*, *paradas por faixa* e *grau de saturação* apresentam menor amplitude e estão posicionados mais acima para as colisões traseiras do que para os RFV em geral. Por exemplo, os 1º e 3º quartis do *fluxo por faixa* para as ocorrências longitudinais foram, respectivamente, de 300 e 450 vphpf; enquanto que para os RFV foram de 180 e 410 vphpf. O intervalo interquartil para *grau de saturação* varia entre 60% e 70% para os acidentes traseiros, já para os RFV, entre 25% e 65%. Esses são indicativos que corroboram com a hipótese 3, de que esse tipo de colisão é mais frequente em RFV de maiores demandas veiculares.

Com relação às colisões transversais e às outras variáveis das colisões traseiras, verifica-se que não há consideráveis diferenças entre os valores dos RFV gerais e com acidente, com exceção ao *ciclo por hora* que será melhor explorado nas análises bivariadas.

Figura 26 - Boxplot RFV associados aos acidentes por tipo





### **4.2.3 Consolidação do banco de dados de acidentes associados às entidades-anuais**

As características de tráfego das entidades-anuais foram coletadas a partir da média (*velocidade, grau de saturação, ciclos por hora e relação fluxo principal-secundária*) e do somatório (*fluxo por faixa e paradas por faixa*) das informações de 15 minutos de cada entidade. Para que uma interseção semaforizada fosse considerada no estudo essa deveria conter informação de, pelo menos, um dia completo no período de 2015 a 2017, sendo calculado o valor da variável a partir da média horária dos atributos diários. Das 95 interseções avaliadas nos RFV, 94 apresentaram informações completas para, no mínimo, um dia completo. Os atributos fixos das entidades foram os mesmos utilizados nos RFV. Os acidentes de trânsito foram coletados para os três anos de estudo a partir de um raio de 50m de cada interseção. A Tabela 8 traz a estatística descritiva das variáveis coletadas para as entidades-anuais.

As estatísticas da Tabela 8 mostram que há uma diversidade de características dentre as 94 interseções semaforizadas estudadas. Por exemplo, o *fluxo por faixa médio anual* (FFMA) oscilou entre 115 e 367 vphpf, indicando condições médias substancialmente diferentes entre as interseções semaforizadas avaliadas. A considerável amplitude da *velocidade* e do *grau de saturação* também revelam a presença de interseções com condições médias de diferentes demandas veiculares. A *relação fluxo via principal-secundária* exhibe a existência de entidades com fluxo homogeneamente dividido entre as direções e de entidades com concentração da demanda em um dos cursos. Assim como nos RFV, apenas quatro interseções semaforizadas possuíam *fiscalização eletrônica* em um raio de 100m, possivelmente inviabilizando a avaliação do efeito dessa variável na frequência e severidade dos acidentes. Nota-se uma concentração de interseções com quatro a seis *faixas de tráfego* e com duas a três *aproximações semafóricas*. Por fim, destaca-se que, em média, 255 dias foram utilizados em cada interseção para obtenção das variáveis médias diárias, com um valor mínimo de sete dias e máximo de 835 dias.

### **4.3 Definição dos agrupamentos de regimes de fluxo veicular**

O banco de dados de RFV contém 4,1M de observações, das quais apenas 668 foram associadas a intervalos com acidentes de trânsito. A grande proporção de RFV sem a presença de acidentes impossibilitou que – aplicando técnicas estatísticas clássicas – fosse estimada a relação entre as características dos RFV e a frequência e severidade dos acidentes. Assim,

visando viabilizar o processo de modelagem, nesta seção as 4,1M observações foram segmentadas a partir da homogeneidade de seus atributos, formando os agrupamentos de regimes de fluxo veicular (ARFV). Esta seção está dividida em três blocos principais: (i) processo de amostragem segmentada dos RFV; (ii) definição do número de agrupamentos; (iii) composição de diferentes configurações de ARFV para aplicação da modelagem. De forma complementar, apresenta-se uma avaliação do impacto nas características dos RFV com a aplicação da correção média nos horários reportados dos acidentes.

Tabela 8 – Estatística descritiva das entidades-anuais

Variável		Média	Desvio padrão	Mín	25° Perc	75° Perc	Máx
Acidentes 2015 (N = 491)	Danos materiais	3,5	3,1	0	2	4	20
	Vítimas	1,7	1,8	0	0	3	10
Acidentes 2016 (N = 497)	Danos materiais	3,3	2,7	0	1	4	12
	Vítimas	2,0	1,9	0	0	3	7
Acidentes 2017 (N = 407)	Danos materiais	2,7	2,3	0	1	4	16
	Vítimas	1,6	1,6	0	0	2	8
Acidentes 2015 a 2017	Danos materiais	9,5	6,5	1	5	12	48
	Vítimas	5,3	3,7	0	3	7	18
Fluxo por faixa médio anual (vei/fx/h)		227	48	115	194	262	367
Velocidade (km/h)		17	4	7	14	20	28
Grau de saturação (%)		44	13	16	35	49	85
Número de paradas por faixa horário médio anual (par/fx/h)		146	43	47	119	172	271
Ciclos semafóricos por hora		39	8	28	32	44	74
Relação fluxo via principal-secundária		2,3	1,2	1,0	1,3	2,9	6,2
Fiscalização eletrônica*		0,05	0,27	0	0	0	2
Número de faixas		5,3	1,6	4	4	6	10
Número de aproximações		2,5	0,6	2	2	3	5
Número de conversões permitidas		2,5	0,7	1	2	3	6
Dias completos disponíveis por interseção		255	212	7	75	432	835

\*Das 94 entidades, 4 possuíam fiscalização eletrônica em raio de 100m

Fonte: Elaborada pelo autor

As técnicas de agrupamentos podem gerar resultados em infinitas dimensões, número que é definido pela quantidade de variáveis inseridas no processo. Visando facilitar a exposição e a análise dos resultados, decidiu-se por desenvolver os agrupamentos utilizando

apenas duas variáveis (resultados em duas dimensões). A Figura 27 traz o gráfico de dispersão e o índice de correlação entre as seis variáveis de tráfego estudadas. Percebe-se que as variáveis *fluxo por faixa*, *paradas por faixa* e *grau de saturação* apresentaram correlação elevada entre si ( $R > 0,75$ ), assim, decidiu-se por considerar apenas uma delas para definição dos agrupamentos. Outra variável importante, *ciclos por hora* também exibiu considerável correlação com o *fluxo por faixa*. Então, por serem a base da relação fundamental do tráfego, por representarem as características de maior interesse nos estudos de segurança viária e por apresentarem correlação média, as variáveis selecionadas foram o *fluxo por faixa* e a *velocidade*.

Os atributos fixos e os outros variáveis foram incorporados aos agrupamentos de forma categórica. Por exemplo, caso sejam definidos cinco agrupamentos de *fluxo por faixa* e *velocidade*, e deseja-se incluir a variável *grau de saturação* na análise (Figura 28). Primeiramente, segmenta-se o *grau de saturação* em X categorias. Em seguida, relaciona-se cada agrupamento a uma das categorias. Supondo  $X = 2$  e que todas as combinações estão presentes na amostra, obtêm-se dez ARFV. Para a etapa de modelagem, pode-se associar a cada um dos dez ARFV um valor médio (média ou mediana) referente a cada uma das variáveis.

#### ***4.3.1 Amostragem segmentada dos regimes de fluxo veicular***

A técnica de agrupamentos hierárquicos utilizando o método aglomerativo com distâncias quadradas (*ward distance*) aplicada na definição dos ARFV é baseada na distância normalizada entre todos os elementos da amostra. Assim, um tamanho amostral de 4,1M resultaria em uma matriz de tamanho 4,1M\*4,1M, inviabilizando execução das rotinas de agrupamentos. Visto a necessidade de reduzir o tamanho amostral, aplicou-se uma segmentação baseada em percentis das variáveis *fluxo por faixa* e *velocidade*. Um tamanho de 9.000 observações foi considerado o valor de equilíbrio entre o tempo de execução das rotinas e o tamanho amostral (cerca de 30 minutos para os agrupamentos e de cinco horas para os indicadores de seleção do número ótimo de agrupamentos, em um computador *Intel Core i5-7500 CPU @ 3.40GHz*).

Figura 27 – Gráfico de dispersão e índice de correlação entre os atributos variáveis dos RFV

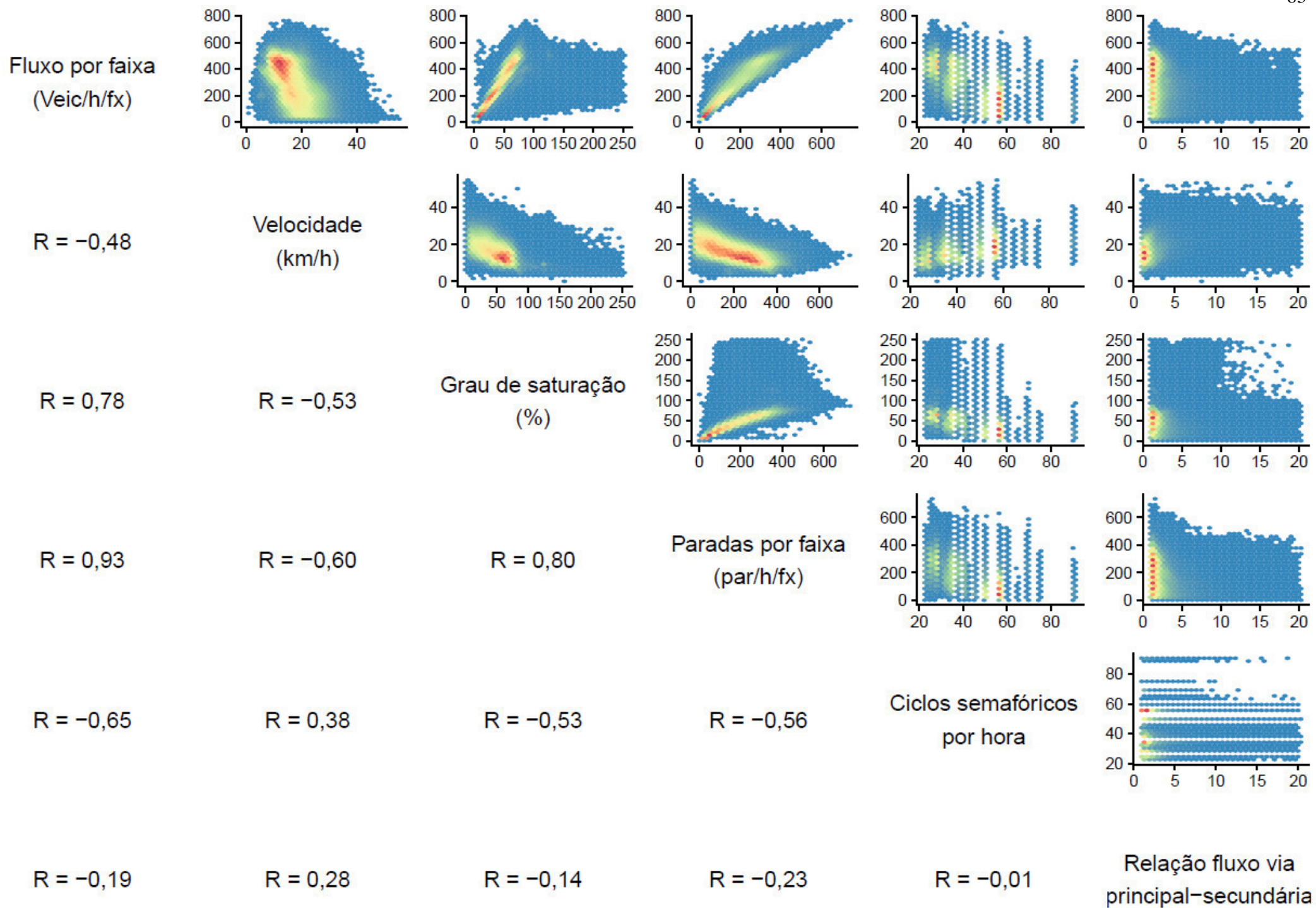
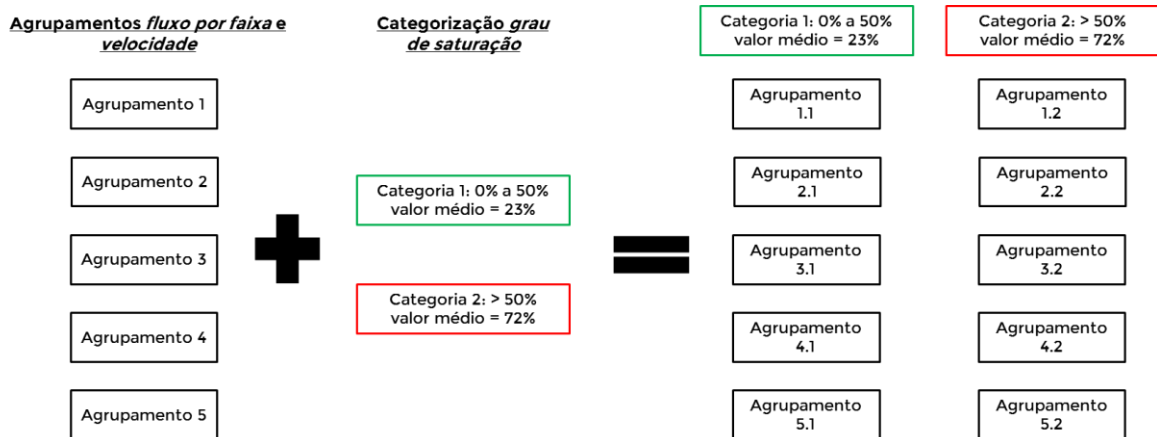


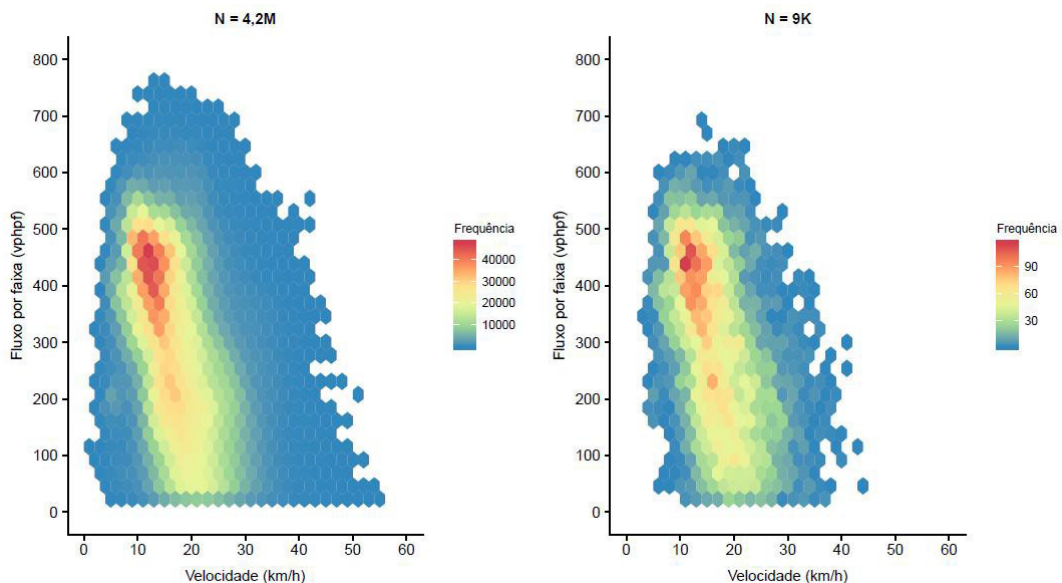
Figura 28 – Exemplo de configuração de agrupamentos de regimes de fluxo veicular



Fonte: Elaborada pelo autor

Visando tentar replicar as distribuições das duas variáveis, os 4,1M RFV foram repartidos em 4.500 percentis de *velocidade* e 4.500 percentis de *fluxo por faixa*, tentando garantir que toda a amplitude das variáveis fosse incorporada na amostra reduzida. A Figura 29 traz a distribuição da amostra completa e da amostra reduzida. Percebe-se que a distribuição das duas amostras é semelhante, com correspondência de densidade e de amplitude em ambos os eixos. Aplicou-se, então, a técnica agrupamentos hierárquicos que computa configurações de agrupamentos variando de 9.000 grupos, em que cada observação representa um agrupamento, a um grupo, caso que todas as observações pertencem ao mesmo grupo.

Figura 29 – Distribuição das amostras completa e reduzida



Fonte: Elaborada pelo autor

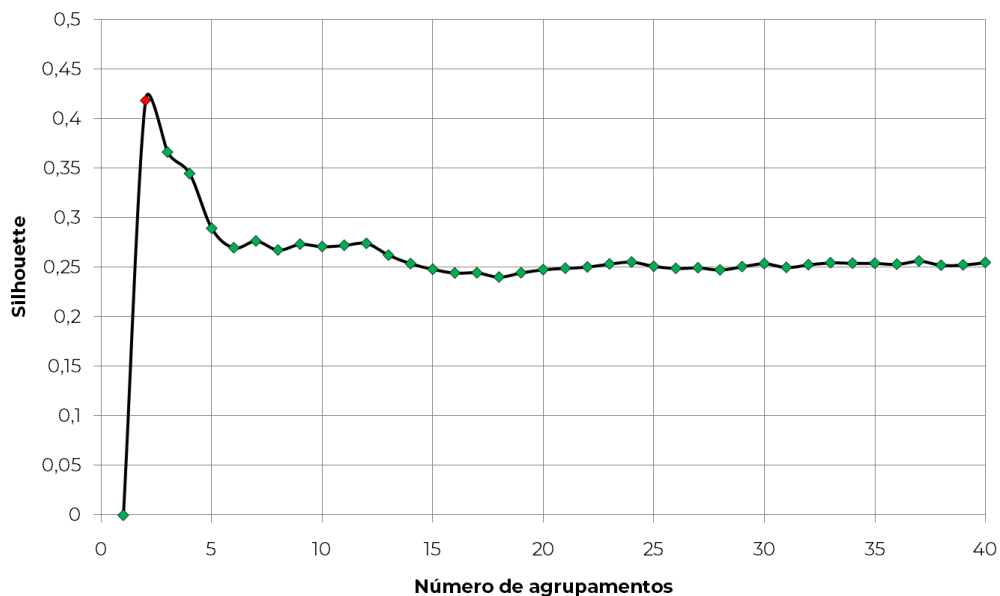
#### 4.3.2 Definição do número de agrupamentos baseados no fluxo por faixa e na velocidade

Das 9.000 configurações possíveis de agrupamentos definidas pelo agrupamento hierárquico, a melhor conformação foi selecionada com o auxílio de três indicadores/testes: (i) *Silhouette test*; (ii) *Gap Statistic test*; (iii) soma dos erros quadráticos entre agrupamentos (SEQA). Após a determinação do número de agrupamentos, apresentou-se uma análise exploratória bivariada entre os agrupamentos e a frequência e severidade dos acidentes.

O *Silhouette test* compara a distância entre o centroide de cada elemento e o centroide de seu agrupamento com a distância entre o centroide de cada elemento e o centroide do agrupamento mais próximo. Em outras palavras, o teste verifica se as características de cada elemento estão fortemente associadas ao seu agrupamento (valor do teste próximo a um) ou se os atributos desse elemento poderiam ser facilmente relacionados a outro grupo (valor próximo a -1). A Figura 30 mostra que para os agrupamentos desenvolvidos para a amostra reduzida, a configuração ótima contém dois agrupamentos ( $K = 2$ ). Verifica-se, ainda, que os agrupamentos com  $K = 3$  e  $K = 4$  estão em uma zona de transição para o valor do teste; que segue praticamente constante a partir de  $K = 5$ .

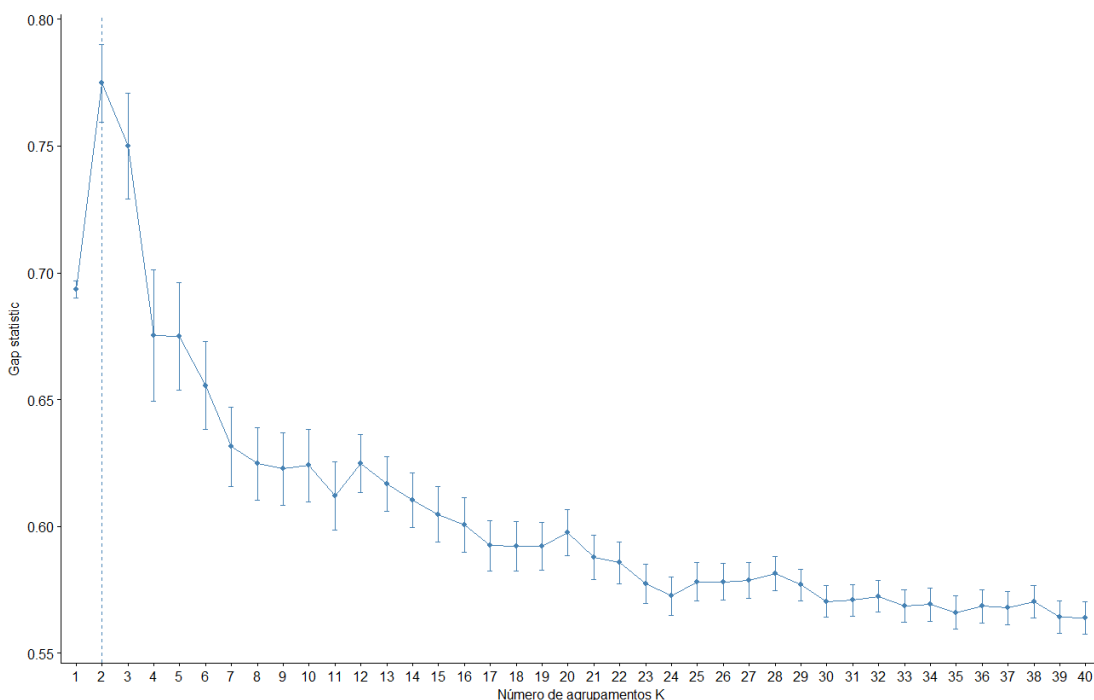
Em geral, a literatura traz que valores do teste abaixo de 0,25 indicam que nenhuma estrutura substancial foi encontrada entre os agrupamentos e que valores entre 0,26 e 0,50 representam fraca estrutura entre os grupos. Os valores até  $K = 14$  se situam acima da linha de *silhouette* igual a 0,25.

Figura 30 – Definição do número ótimo de agrupamentos: *Silhouette test*



O *Gap Statistic test* compara a soma dos erros entre agrupamentos da amostra utilizada com o somatório dos erros caso fosse aplicada uma amostra com distribuição aleatória (sem estrutura de agrupamentos). Quanto maior o valor do indicador, maior o indício de que a estrutura formada está distante de uma distribuição aleatória. Outro critério para seleção do número de agrupamentos é optar pelo menor valor  $K$  que satisfaz a equação:  $Gap(K) \geq Gap(K+1) - s_{k+1}$ . Assim como no teste anterior, o número ótimo de agrupamentos sugerido pelo método é  $K = 2$  (Figura 31).

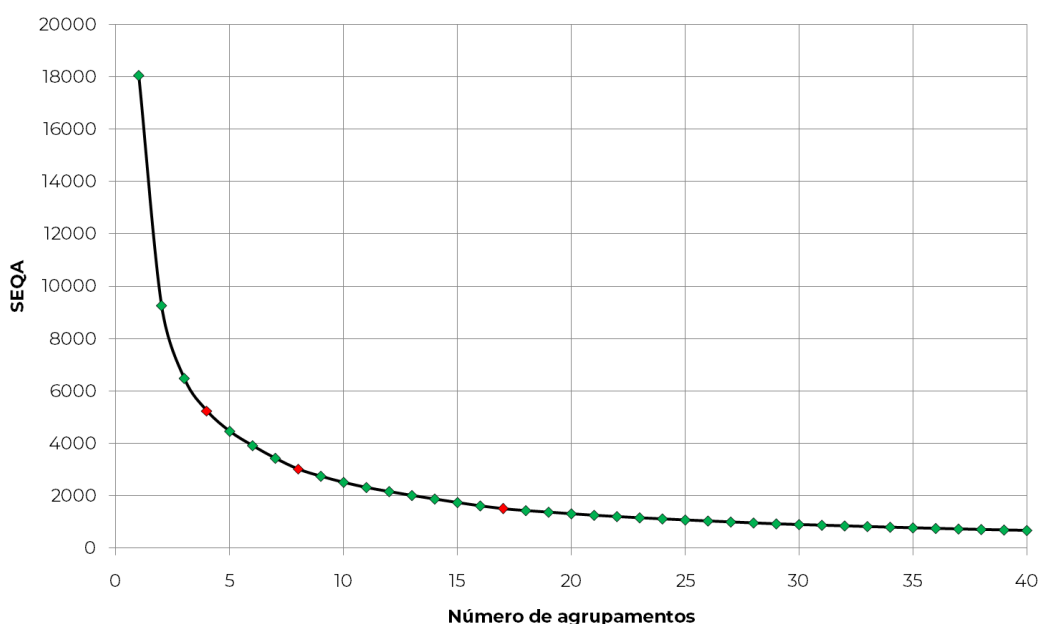
Figura 31 – Definição do número ótimo de agrupamentos: *Gap Statistic test*



Fonte: Elaborada pelo autor

A SEQA, como sugere sua nomenclatura, mede a soma da distância entre cada elemento e o centroide de seu agrupamento. O aumento no número  $K$  traz reduções na SEQA, fazendo com que o indicador seja aplicado na identificação de até quando essas retrações nos erros são significativas. Pelo *elbow method* verifica-se que o “cotovelo” da curva está entre  $K = 4$  e  $K = 8$  (Figura 32). Utilizando a porcentagem de redução da SEQA entre agrupamentos sucessivos, por exemplo: redução na SEQA de 19% quando há aumento de  $K$  de três para quatro, as configurações com maior número de agrupamentos que resultaram em reduções máximas de 15%, 10% e 5% foram, respectivamente,  $K = 4$ ,  $K = 8$  e  $K = 17$  (destacados em vermelho na Figura 32).

Figura 32 – Definição do número ótimo de agrupamentos: soma dos erros quadrados entre agrupamentos



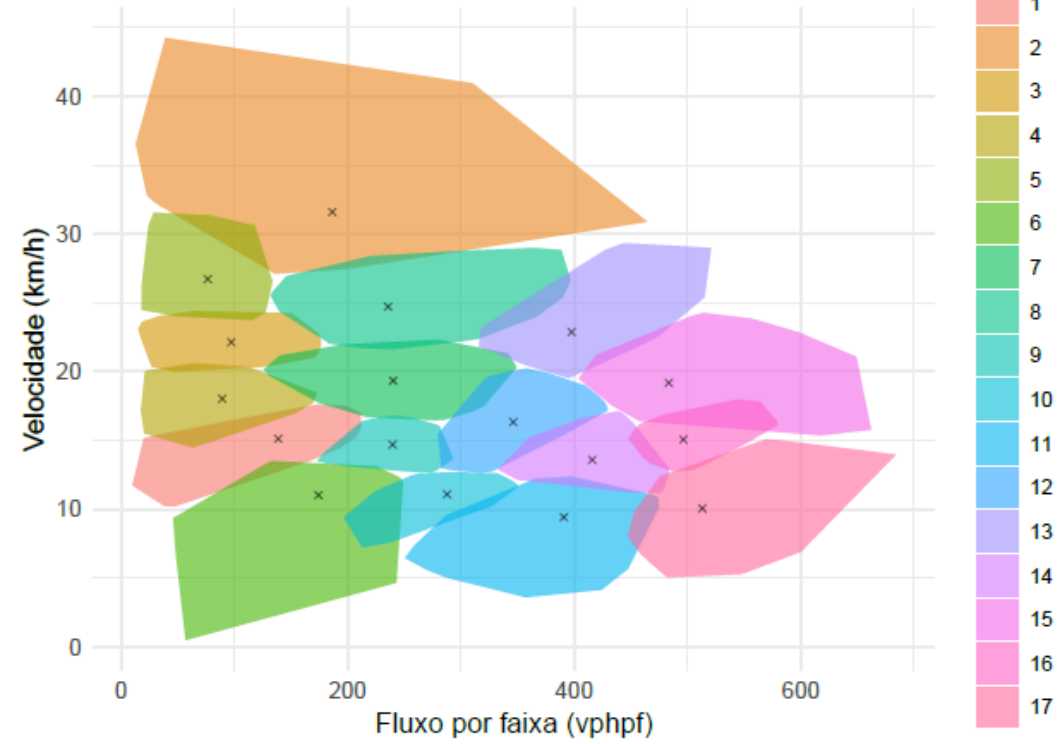
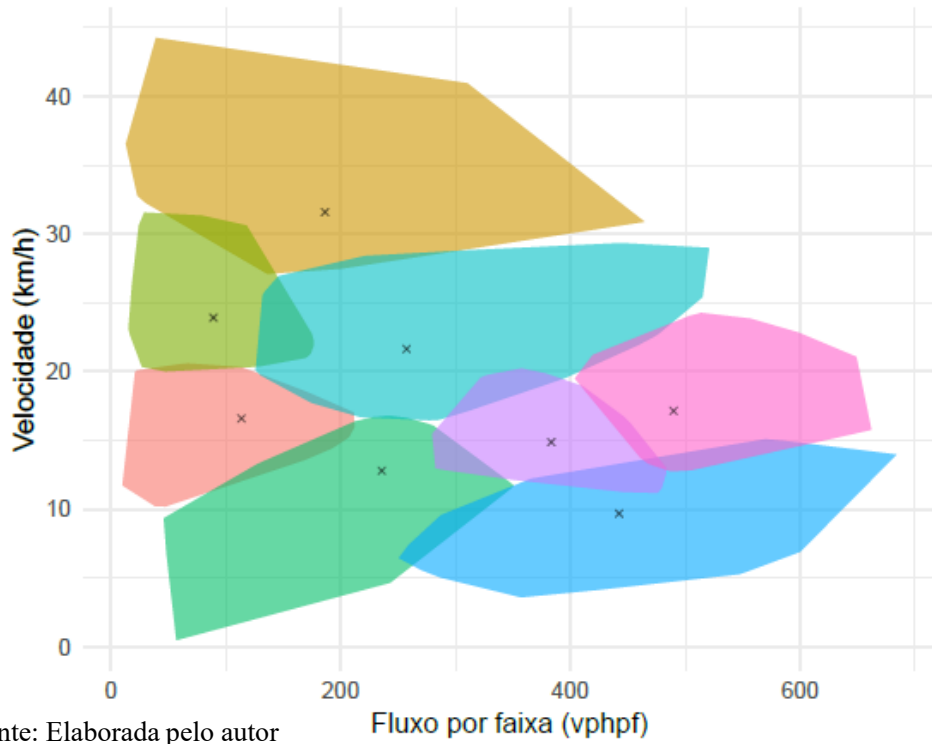
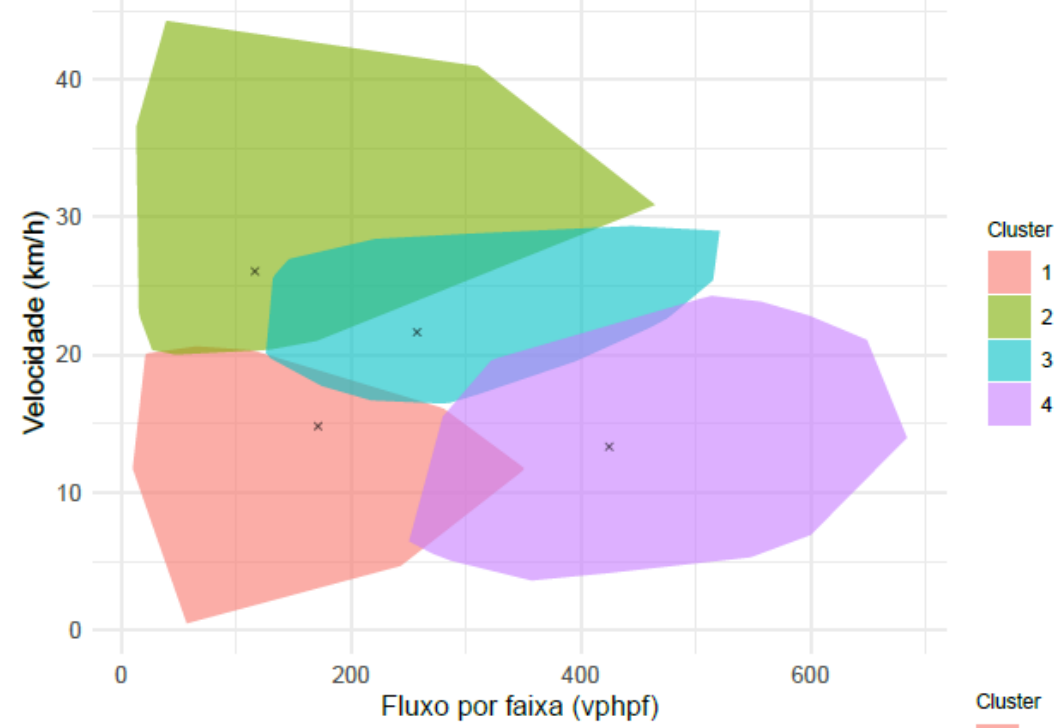
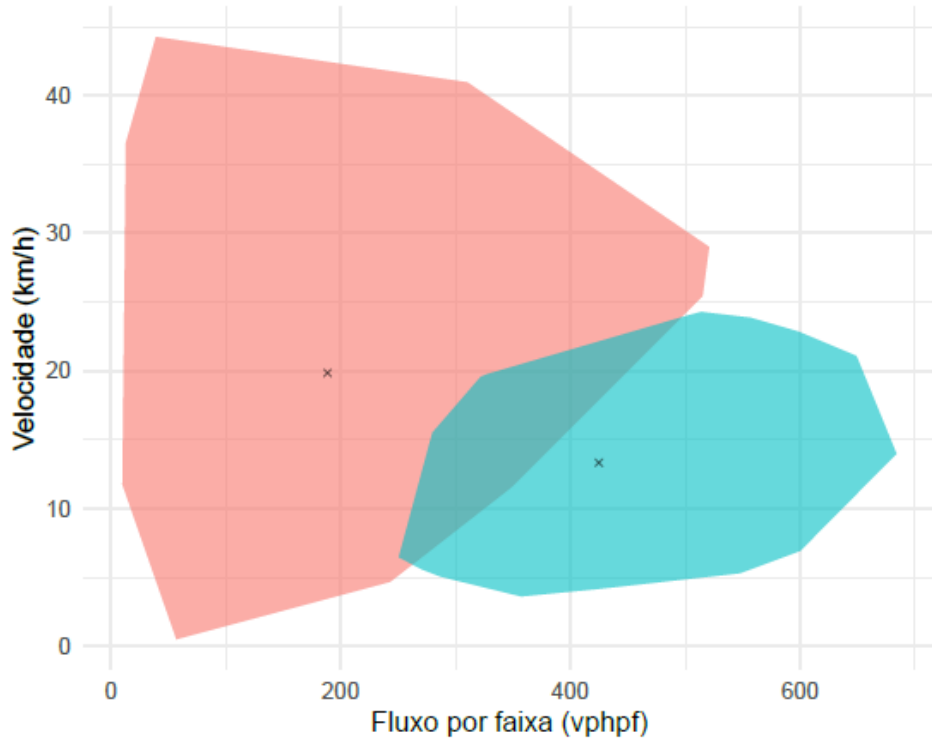
Fonte: Elaborada pelo autor

A Figura 33 apresenta as quatro configurações indicadas com a aplicação dos indicadores. A primeira conformação com dois agrupamentos – sugerida pelos dois primeiros testes – produziu conjuntos que tenderam a representar os regimes congestionados e não congestionados. Embora represente a configuração mais adequada para o banco de dados avaliado, a utilização de apenas dois agrupamentos vai de encontro à abordagem desagregada que pretende ser adotada neste estudo, assim, decidiu-se por avaliar as configurações recomendadas pelo indicador SEQA.

A configuração com  $K = 4$  segregou o agrupamento de regime não congestionado caracterizado em  $K = 2$  em outros três agrupamentos. Ainda assim, o nível de desagregação não aparenta ser o mais indicado para condução das análises. Por exemplo, não há uma diferenciação nos níveis de velocidade no agrupamento que reflete o regime congestionado (grupo quatro), ou seja, velocidades entre 5 km/h e 20 km/h seriam tratadas como similares. O mesmo vale para o grupo um, representando um regime com baixos *fluxo por faixa* e *velocidade*.

Surgiu, então, a decisão entre  $K = 8$  e  $K = 17$  que foi baseada em dois aspectos: (i) análise visual da conformação dos grupos; (ii) relação entre a quantidade de agrupamentos e de acidentes disponíveis para análise. Devido às poucas observações com velocidades acima de 30 km/h, as duas configurações não segregaram o grupo de maior velocidade.





A conformação com  $K = 8$  mantém alguns agrupamentos (grupos 4, 5 e 6) com considerável amplitude – principalmente no eixo das abscissas – que pode afetar na qualidade dos modelos. Como esperado, a composição com  $K = 17$  apresenta agrupamentos mais concisos, permitindo a formação de grupos de menores amplitudes, possibilitando um maior grau de detalhamento que pode ser importante na estimação da relação entre as variáveis e os acidentes de trânsito.

A discussão anterior sugere que a conformação com  $K = 17$  é a mais adequada para prosseguimento das análises, entretanto o tamanho da amostra dos acidentes pode alterar o panorama. IMPRIALOU et al. (2016) associaram 9.310 acidentes a 2.400 ARFV, representando uma relação acidente-agrupamento de 3,87. A amostra de acidentes contém 668 ocorrências, sendo 432 com danos materiais e 236 com vítimas. Por exemplo, uma configuração de ARFV incluindo as variáveis *número de faixas* e *ciclos por hora* segregadas em três e quatro categorias, respectivamente, resultaria em 204 agrupamentos ( $17*3*4$ ), o que representa uma relação acidente com vítima-agrupamento de 1,16. Assim, seria observada uma alta frequência de agrupamentos com zero acidente, possivelmente demandando a aplicação de modelos inflados de zero que – como citado na seção de revisão – possuem premissas que podem não ser adequadas para explicação do fenômeno de acidentes. Dessa forma, decidiu-se aplicar a seguinte estratégia: (i) prosseguimento das análises exploratórias e da modelagem com a conformação  $K = 8$ ; (ii) caso o número de agrupamentos do modelo de melhor ajuste permita a utilização de  $K = 17$  (i.e., possua poucas variáveis explicativas), o modelo também será calibrado para essa configuração, sendo realizada uma comparação com o modelo anterior.

A Tabela 9 traz um resumo das estatísticas dos oito ARFV. Percebe-se que a quantidade de elementos oscila por agrupamento, indicando que a frequência de acidentes observada por agrupamento deve ser expressa considerando o fator exposição. Para evitar elevadas quantias de casas decimais, decidiu-se por trabalhar utilizando a unidade de número de acidentes por 10.000 horas de RFV. A Figura 34 traz os acidentes por severidade de cada ARFV.

O gráfico dos acidentes com danos materiais mostra que regimes com elevados *fluxos por faixa* e baixas *velocidades* apresentaram maior probabilidade de ocorrências desse tipo do que em regimes menos congestionados. Em termos numéricos, em RFV com *velocidade* média de 9,7 km/h e *fluxo por faixa* médio de 442 vphpf ocorreram cerca de seis acidentes com danos materiais a cada 10.000 horas, já em RFV com *velocidade* média de 23,9 km/h e

*fluxo por faixa* médio de 89 vphpf o valor caiu para dois acidentes, corroborando com a hipótese 1.

Tabela 9 – Resumo dos ARFV com K = 8

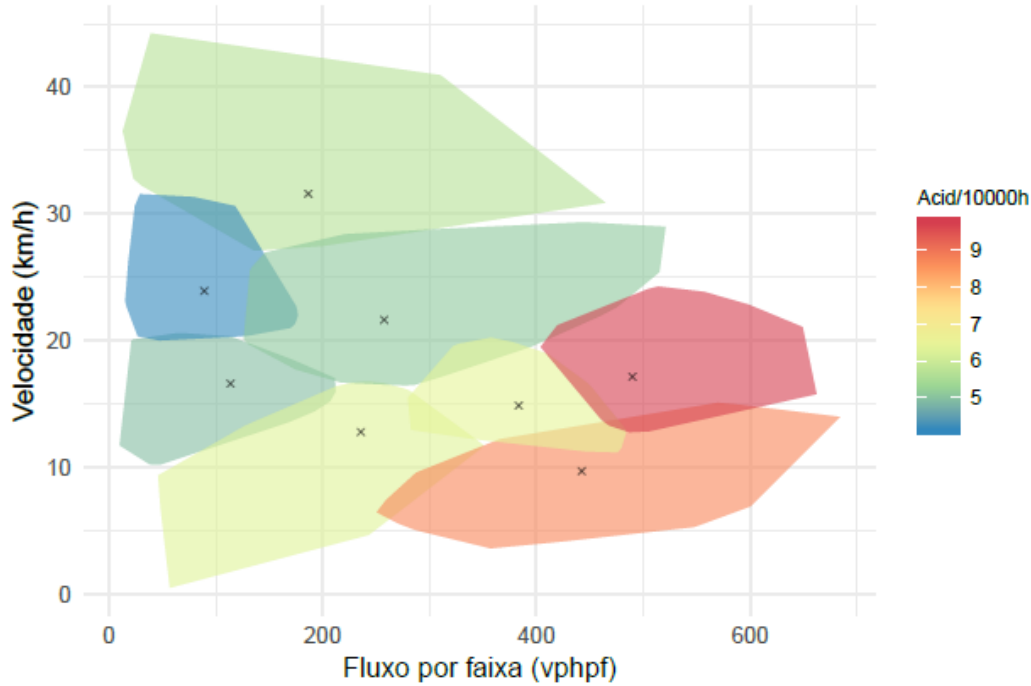
<b>Agrupamento</b>	<b>Velocidade</b>	<b>Fluxo por faixa</b>	<b>Elementos</b>	<b>% de elementos</b>
1	16,6	114	527.738	12,9%
2	31,6	186	190.635	4,6%
3	23,9	89	414.491	10,1%
4	12,8	236	533.649	13,0%
5	21,6	257	596.878	14,5%
6	9,7	442	751.663	18,3%
7	14,9	383	668.526	16,3%
8	17,1	490	418.987	10,2%

Fonte: Elaborada pelo autor

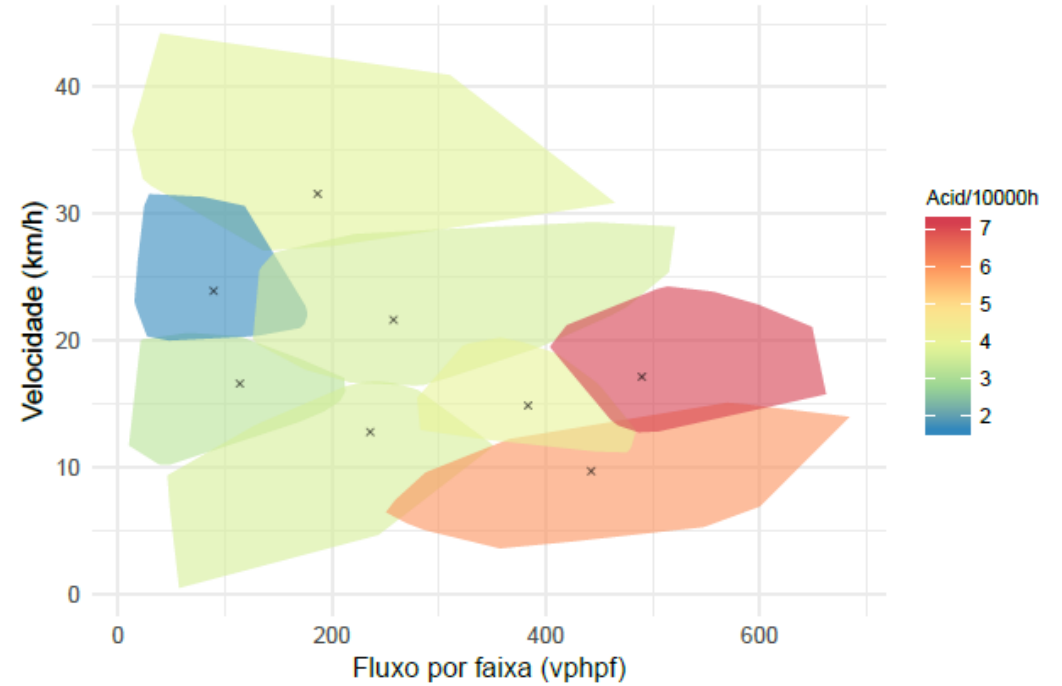
O gráfico dos acidentes com vítimas apresenta que os regimes mais críticos estão distribuídos na amplitude do *fluxo por faixa* e da *velocidade*. Em comparação aos acidentes com danos materiais, percebeu-se um deslocamento da ocorrência dos acidentes para os regimes menos congestionados, mantendo, também, os incidentes em regimes mais demandados. Em resumo: não foi possível verificar – visualmente – tendências entre as duas variáveis e os acidentes com vítimas.

Algumas hipóteses podem ser levantadas para tal: (i) como a maioria dos acidentes com vítimas ocorreu com usuários vulneráveis (73%), esses acidentes também ocorreram em RFV de baixas *velocidades médias*, visto que um atropelamento ou uma colisão envolvendo motociclistas ou ciclistas possivelmente produziria uma vítima se o impacto ocorresse a 25 km/h ou a 50 km/h (relembra-se que a velocidade estimada pelo *SCOOT* considera o tempo em que o veículo ficou parado); (ii) apenas uma parcela dos acidentes com vítimas foi explicada pelo desempenho de velocidade pelos condutores: em regimes de menor demanda veicular – consequentemente de menor exposição aos acidentes – o desenvolvimento de velocidade foi preponderante para ocorrência de acidentes com vítimas; (iii) o método de estimação da velocidade pelo *SCOOT* não foi capaz de identificar veículos que conseguem desempenhar altas velocidades mesmo em regimes congestionados, motocicletas principalmente.

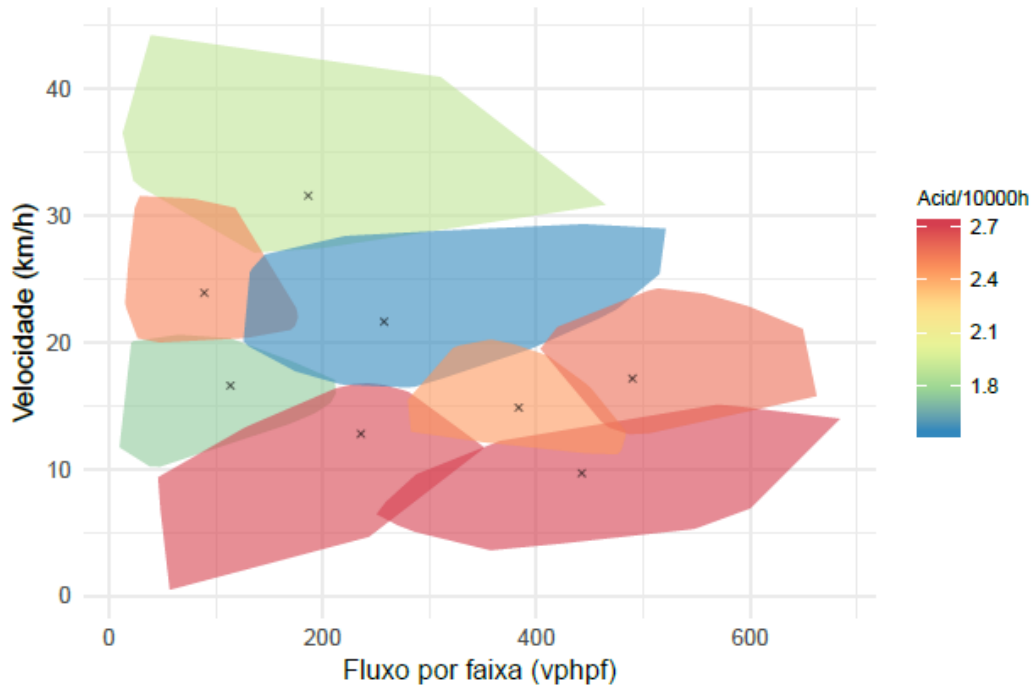
**Acidentes totais (N = 668)**



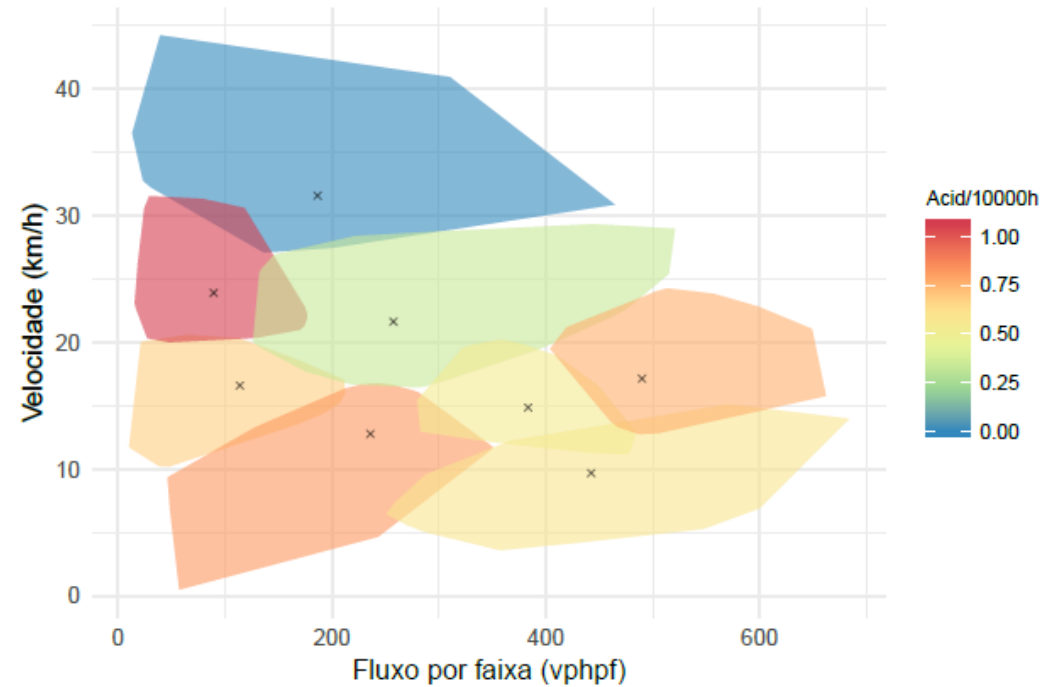
**Acidentes com danos materiais (N = 432)**



**Acidentes com vítimas (N = 236)**



**Acidentes com vítimas - excluindo com motos e pedestres (N = 64)**



O gráfico dos acidentes com vítimas não vulneráveis corrobora as hipóteses (i) e (ii) levantadas acima, visto que o deslocamento dos pontos mais críticos para regimes menos congestionados sugere maior importância da *velocidade* para ocorrência de vítimas em veículos protegidos. Isso era esperado, uma vez que impactos a baixas velocidades tendem a gerar apenas danos materiais nesses usuários.

Destaca-se, ainda, a importância que outras variáveis, como *grau de saturação*, *ciclos por hora* e *número de faixas*, podem ter no entendimento do fenômeno dos acidentes. Essas variáveis foram incorporadas às configurações de ARFV na próxima seção.

Além disso, outros fatores de difícil mensuração, por exemplo *avanços semaforicos*, *ingestão de bebida e direção*, *uso do cinto de segurança e do capacete*, entre outros; também são relevantes para explicação da ocorrência dos acidentes – principalmente dos com vítimas – mas não foram abordados neste trabalho.

#### **4.3.3 Definição das configurações dos agrupamentos de regimes de fluxo veicular**

O objetivo desta seção é a definição das configurações de ARFV a serem exploradas na etapa de modelagem. Para isso, os atributos fixos e as variáveis de tráfego não incorporadas aos agrupamentos foram categorizados e selecionados para as composições dos grupos. Além disso, apresentou-se a análise exploratória multivariada de algumas das variáveis.

A partir dos percentis apresentados na Tabela 10 determinou-se a categorização das variáveis, sendo cada categoria associada a um valor médio da variável. As variáveis contínuas (*paradas por faixa*, *ciclos por hora* e *relação princ-sec*) foram segregadas em quatro categorias baseadas nos valores de quartis, com exceção do *grau de saturação* que se decidiu também incluir os *outliers* superiores como uma categoria. Para as variáveis discretas (*nfx*, *app* e *conversões*) percebeu-se que um mesmo valor da variável correspondia a mais de um quartil, assim, tais foram distinguidas em três ou quatro categorias. A *fiscalização eletrônica* foi categorizada de forma binária, representando presença ou não de equipamento em raio de 100m da interseção.

Definidas as categorias das variáveis, os índices de correlação entre os atributos dos RFV são apresentados na Figura 35. Com relação às características de tráfego, verifica-se que *fluxo por faixa*, *paradas por faixa* e *grau de saturação* apresentaram alta correlação entre si, sendo um indicativo de que essas variáveis não devem ser incorporadas simultaneamente aos modelos, visto que essa correlação dificultaria a interpretação da parcela de variável na

explicação da ocorrência dos acidentes. Devido à quase perfeita correlação entre *fluxo por faixa* e *parada por faixa*, apenas a primeira foi explorada nas análises. Como esperado, nota-se intermediária correlação entre o *ciclos por hora* e o *fluxo por hora*. A inclusão simultânea dessas variáveis demandou atenção aos efeitos nos coeficientes dos modelos.

Tabela 10 – Categorização das variáveis para definição das configurações de ARFV

Categoria	Variável								
	Grau de saturação	Paradas por faixa	Ciclos por hora	Rel. princ-sec	Nfx	App	Conversões	FE	
1	a	<28%	1-96	22-28	1,0-1,2	4	2	1-2	0
	b	17%	53	26	1,1	4	2	2,0	0
2	a	29%-46%	97-180	30-37	1,2-1,6	5-6	3	3	1 ou 2
	b	37%	135	34	1,3	5,7	3	3	1
3	a	47%-62%	181-273	40-50	1,6-2,3	7-8	4-5	4-6	-
	b	55%	223	46	1,9	7,6	4,16	5,5	-
4	a	63%-110%	274-728	56-90	2,3-20	10	-	-	-
	b	73%	351	58	4,1	10	-	-	-
5*	a	>110%	-	-	-	-	-	-	-
	b	150%	-	-	-	-	-	-	-

\*Valores *outliers* superiores. a: intervalo da categoria; b: valor médio da categoria

Fonte: Elaborada pelo autor

Figura 35 – Matriz de correlação entre as variáveis dos RFV

Variáveis	Fluxo por faixa	Velocidade	Grau de saturação	Paradas por faixa	Ciclos por hora	Rel. princ-sec	Nfx	App	Conversões	FE
Fluxo por faixa	-									
Velocidade	-0.48	-								
Grau de saturação	<b>0.78</b>	-0.53	-							
Paradas por faixa	<b>0.93</b>	-0.60	<b>0.80</b>	-						
Ciclos por hora	-0.65	0.38	-0.53	-0.56	-					
Rel. princ-sec	-0.19	0.28	-0.14	-0.23	-0.01	-				
Nfx	0.02	0.00	0.14	0.03	-0.12	0.22	-			
App	0.03	-0.03	0.13	0.04	-0.11	0.21	<b>0.92</b>	-		
Conversões	0.06	-0.03	0.12	0.08	-0.12	0.01	<b>0.76</b>	<b>0.78</b>	-	
FE	0.02	0.12	0.03	0.00	-0.01	-0.04	0.10	0.10	0.01	-

Fonte: Elaborada pelo autor

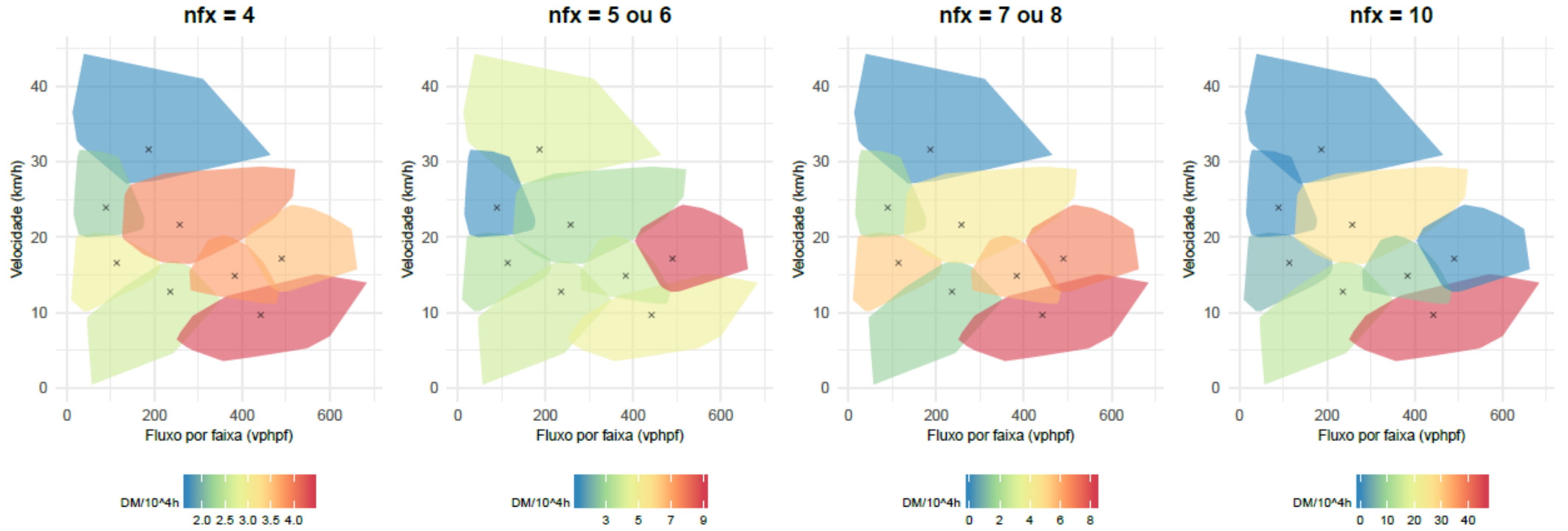
A respeito dos atributos fixos, o *número de faixas*, *de aproximações* e *de conversões* foram altamente correlacionados. Decidiu-se por trabalhar com o *número de faixas*, por representar a variável mais comumente avaliada em outros estudos e por corresponder ao denominador da variável *fluxo por faixa*. As configurações de ARFV, então, contêm as combinações de todas as variáveis, considerando as restrições elencadas acima: *fluxo por faixa*, *velocidade*, *grau de saturação*, *ciclos por hora*, *relação principal-secundária*, *número de faixas* e *fiscalização eletrônica*.

As Figuras 36 e 37 apresentam a frequência relativa de acidentes por agrupamento com a adição das categorias de *número de faixas* e *ciclos por hora*, respectivamente (notar que cada gráfico possui escala diferente). A ordem de grandeza das escalas da Figura 36 mostra que tanto para os acidentes com danos materiais quanto para os com vítimas o aumento do *número de faixas* exibiu relação a um incremento na probabilidade de ocorrência de acidentes, corroborando com a hipótese 4. O *número de faixas* está relacionado à possibilidade de movimentação dos condutores, permitindo manobras laterais, aumentando, assim, a probabilidade de ocorrência de conflitos veiculares. Com a utilização da variável de exposição por quantidade de faixas era esperado que locais com maiores *número de faixas* apresentassem frequências de acidentes mais elevadas, visto que o fluxo veicular é proporcional à quantidade de faixas.

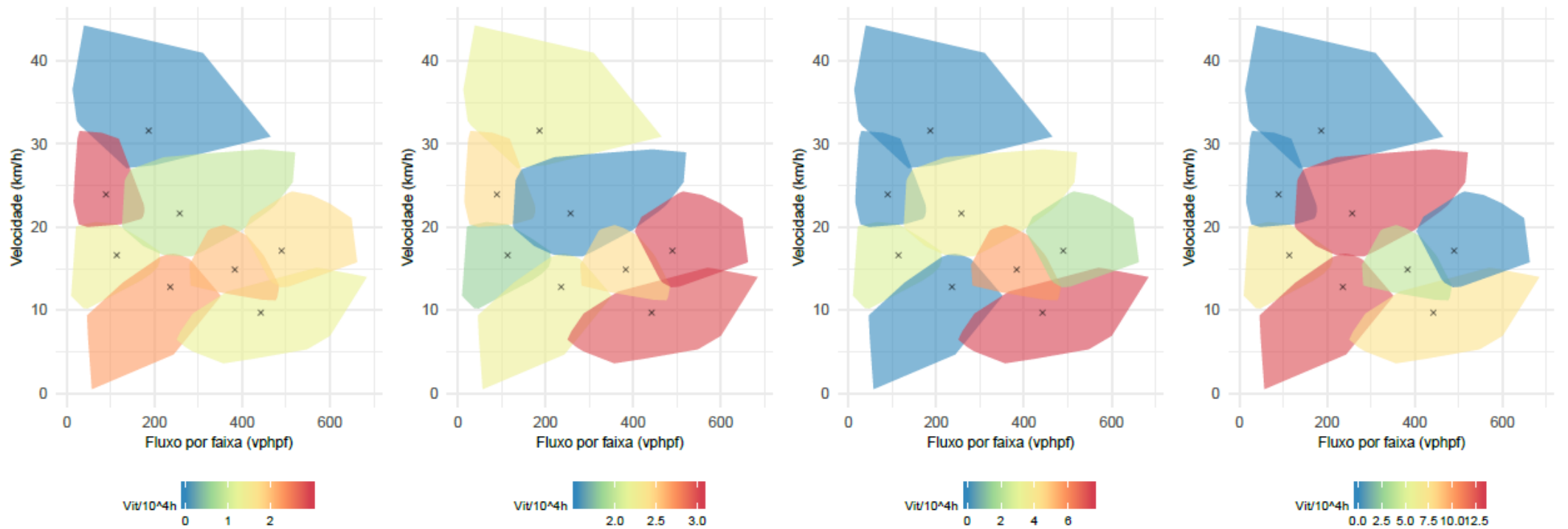
Percebe-se que os RFV com dez faixas de tráfego foram, definitivamente, os que apresentaram maior probabilidade de ocorrência dos dois tipos de acidentes. Para os danos materiais, por exemplo, verifica-se ocorrência de 46 acidentes a cada 10.000 horas no regime de maior congestionamento. A amostra contém apenas três interseções semaforizadas com essa característica: Av. Aguanambi x Av. Domingos Olímpio, Av. Sen. Virgílio Távora x Av. Abolição e Av. Aguanambi x Av. Soriano Albuquerque. Essas interseções estão entre as de maior porte da cidade, sendo um indicativo de que entidades com essas características devem ser constantemente avaliadas por gestores das cidades, além de serem potenciais candidatas a receber fiscalização eletrônica ou outras medidas em benefício da segurança viária.

A Figura 37 apresenta a relação entre a frequência de acidentes relativa e os ARFV categorizados pelo número de *ciclos por hora*. A presença do semáforo e, conseqüentemente, da variável *ciclos por hora* traz um importante componente para o fenômeno dos acidentes: a exposição ao conflito veicular longitudinal. Por exemplo, em um trecho sem presença de semáforo espera-se que os conflitos entre veículos sejam resultados principalmente da interação por deslocamentos laterais ou por movimentos transversais de entrada ou saída de veículos/pedestres na via. A inserção do semáforo traz constantes interações longitudinais

Acidentes com danos materiais

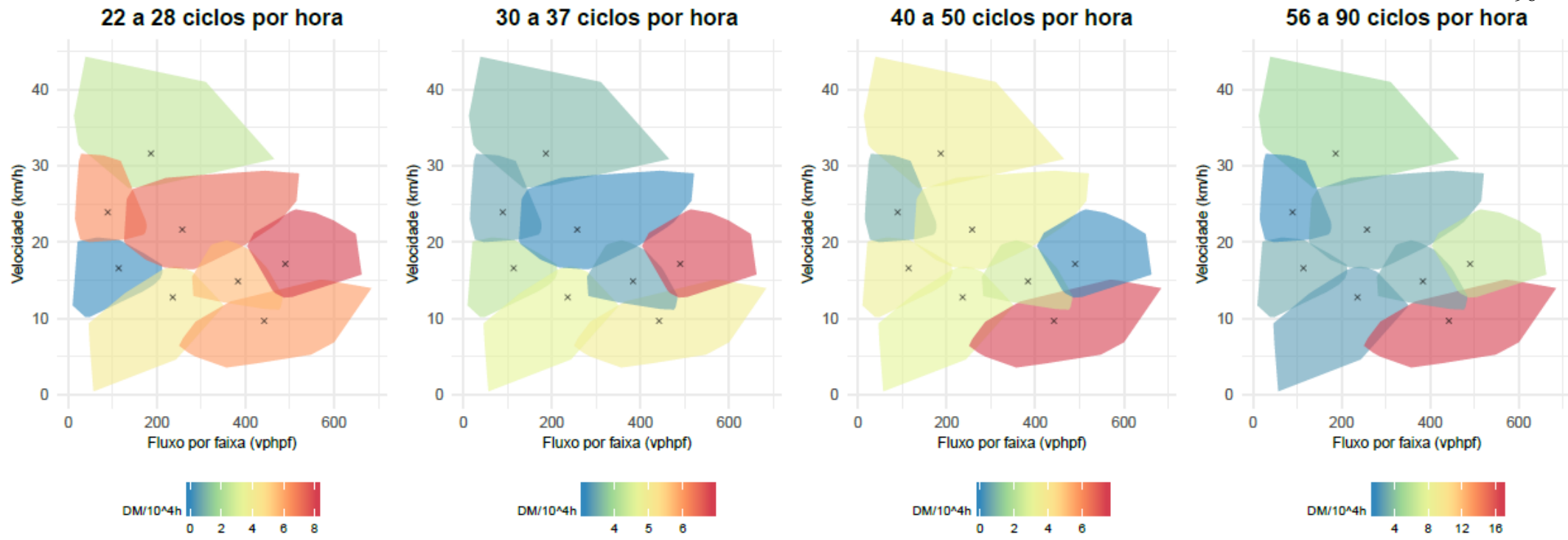


Acidentes com vítimas

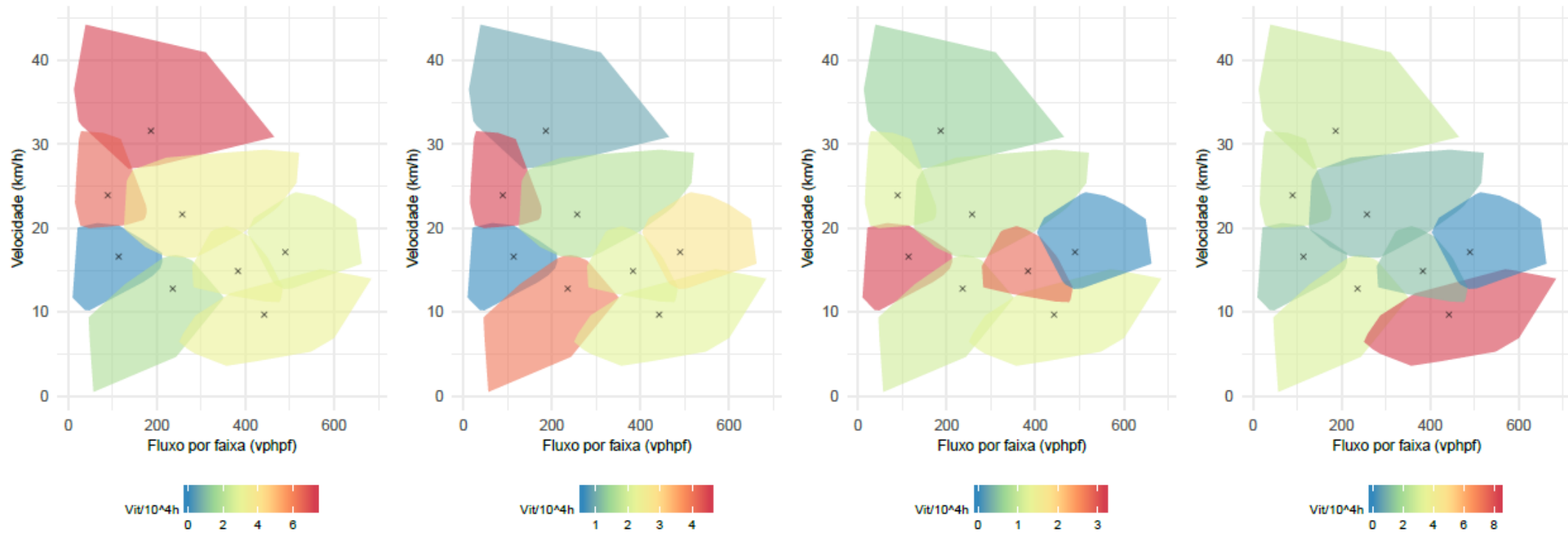




Acidentes com danos materiais



Acidentes com vítimas



entre os veículos, conjecturando-se que à medida que o número de ciclos semaforicos em um regime cresce, aumenta também a exposição às colisões traseiras. Além disso, supõe-se que ciclos semaforicos mais longos permitem, principalmente em casos em que os semaforos sejam coordenados, o desempenho de altas velocidades por uma maior parcela de veículos, representando regimes situados nos agrupamentos com velocidades médias superiores a 25 km/h .

A avaliação da ordem de grandeza dos gráficos dos dois tipos de acidentes mostra que os ARFV com maior número de *ciclos por hora* foram os que apresentaram maior probabilidade de ocorrência de acidentes, principalmente nos regimes de alta demanda veicular e baixa *velocidade*. Chamam a atenção os expressivos valores de 16 e 8 acidentes com danos materiais e vítimas, respectivamente, a cada 10.000 horas, corroborando com a hipótese 6 de que o aumento do número de *ciclos por hora* está relacionado a maiores frequências de acidentes.

Verifica-se nos gráficos de acidentes com vítimas certo deslocamento dos ARFV mais críticos com o aumento do *ciclos por hora*: em regimes de ciclos semaforicos mais longos houve uma tendência dos acidentes serem mais frequentes em situações de menor demanda veicular, possivelmente devido ao desenvolvimento de maiores *velocidades* possibilitado pelo fluxo de maior continuidade. Já para os regimes de ciclos semaforicos mais curtos, os ARFV mais críticos foram aqueles de maior *fluxo por faixa*, indicando que o movimento de parada-arranque também possui relevância na frequência de acidentes com vítimas. O aumento no número de *ciclos por hora* também pode estar relacionado ao incremento da movimentação de motociclistas nos corredores formados entre os veículos, chamados de corredores virtuais, visto que os condutores geralmente buscam se aproximar da faixa de retenção para facilitar a movimentação na mudança do estágio semaforico.

Comparando esses gráficos com a imagem dos acidentes com vítimas nos ARFV sem a segregação por *ciclos por hora* (Figura 34), percebe-se a significância dessa variável para o fenômeno. Em termos numéricos, o ARFV de maior *velocidade* média apresentou cerca de dois acidentes com vítimas por 10.000 horas na Figura 34, já considerando apenas os ARFV com 22 a 28 *ciclos por hora* esse número sobe para aproximadamente seis (Figura 37). O mesmo vale para o ARFV que caracteriza maiores demandas veiculares e baixas *velocidades*: a categoria de 50 a 90 *ciclos por hora* possui indicador próximo a oito, enquanto o ARFV geral foi de 2,7. Isso é um forte indicativo de que as combinações (i) *baixa demanda veicular* + *ciclos semaforicos longos* e (ii) *alta demanda veicular* + *ciclos semaforicos curtos* foram as

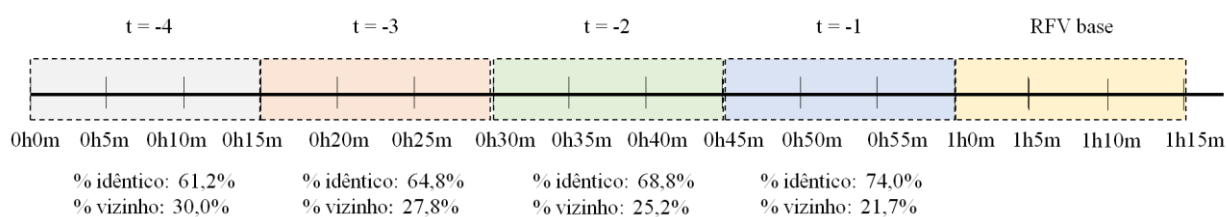
mais críticas em termos de ocorrência de acidentes com vítimas. Comportamento similar também foi verificado nos gráficos de acidentes com danos materiais.

#### **4.3.4 Análise complementar I: avaliação do impacto nas características dos RFV com a aplicação da correção média nos horários reportados dos acidentes**

O objetivo desta análise complementar é verificar o impacto nas características dos RFV com a aplicação da correção média nos horários reportados dos acidentes (objetivo I). Em outras palavras esta seção pretende responder ao seguinte questionamento: supondo que um acidente com vítima ocorreu (e foi reportado) no horário de 10h25m, a correção média aplicada de 44 minutos – definida na seção um – faria com o que o novo horário fosse de 9h41m. O intervalo precursor do primeiro horário é de 10h00m-10h15m, já o do segundo é de 9h15m-9h30m. As características de RFV do segundo intervalo são consideravelmente diferentes quando comparadas as do primeiro intervalo?

Para desenvolver a análise foram selecionados aleatoriamente 50.000 RFV definidos como “RFV base”. Para cada RFV base foram obtidos os quatro RFV temporalmente anteriores, considerando a mesma interseção semaforizada do RFV base. A comparação se deu entre o agrupamento atrelado ao RFV base e aos RFV anteriores, verificando se eles são idênticos ou vizinhos. Caso os RFV estejam situados no mesmo agrupamento, é um indicativo de que o retrocesso temporal não está impactando as características de tráfego no nível de agregação estudado. Caso o agrupamento associado ao RFV anterior seja vizinho ao grupo do RFV base, é um indício de que o regresso temporal está impactando moderadamente as características de tráfego. A Figura 38 traz um resumo dos resultados encontrados.

Figura 38 – Análise complementar I: impacto do retrocesso temporal nas características de tráfego



Fonte: Elaborada pelo autor

Os resultados mostram que o aumento do afastamento temporal tende a gerar menores porcentagens de agrupamentos idênticos. Tomando como exemplo um acidente com vítima

reportado no horário de 1h25m, caso o horário reportado seja o de real ocorrência do acidente, seu intervalo precursor seria o mostrado como RFV base na Figura 38. A aplicação da correção média de 44 minutos, resultaria em um horário teórico de ocorrência de 0h41m; em que o intervalo precursor corresponderia à caixa “ $t = -3$ ”. Nesse cenário há uma chance de 64,8% do agrupamento associado ao acidente ser idêntico ao grupo realmente precursor e uma probabilidade de 27,8% desses agrupamentos serem vizinhos.

Entretanto, em poucos casos os acidentes foram reportados próximos aos horários da detecção da perturbação na velocidade, diferença cujo valor médio foi de 15 minutos para os acidentes com vítimas (Tabela 3). Assim, em um cenário em que o acidente exemplificado acima segue o comportamento médio, seu horário mais provável de ocorrência seria de 1h10m, fazendo com que a diferença entre o RFV base e o RFV com a correção média esteja em “ $t = -2$ ”, aumentando as probabilidades de obtenção de agrupamentos idênticos ou vizinhos.

Em termos de acidentes com danos materiais, a correção média adotada foi de 66 minutos, enquanto a diferença média entre os horários reportados e de detecção de perturbação na velocidade foi de 21 minutos. No geral, considera-se que o método adotado para obtenção de condições precursoras sem interferência do acidente nas variáveis de tráfego é satisfatório, visto que esse está atendendo suas principais demandas: (i) obtenção de condições sem influência do acidente; (ii) baixo impacto nas características de tráfego com o regresso temporal.

#### **4.4 Modelagem estatística: relação entre os acidentes e as variáveis de tráfego, de operação e de geometria**

A modelagem aplicando a distribuição de Poisson demanda que algumas premissas sejam verificadas: (i) a variável dependente deve ser discreta, inteira e não negativa; (ii) as observações devem ser independentes; (iii) a distribuição da variável dependente deve seguir Poisson, apresentando valores para média e variância próximos. A utilização de dados de acidentes, ou seja, valores inteiros não negativos garantiu a primeira premissa. A segunda premissa foi assegurada partindo da hipótese de que a ocorrência de um acidente em um raio de 50m de uma interseção semaforizada não altera a probabilidade de acontecimento de acidente em outra interseção. A última premissa foi testada com aplicação do indicador  $\sigma_d$  que mede a aderência do modelo aos dados utilizados. As premissas para a distribuição Binomial Negativa são similares às de Poisson. Nesta seção são apresentadas as modelagens

considerando as abordagens por ARFV e por entidades, exibindo, em seguida, uma comparação entre os resultados obtidos.

#### **4.4.1 Modelagem estatística: abordagem por agrupamentos de regimes de fluxo veicular**

Como destacado na etapa do método, o processo de seleção das variáveis foi realizado com uma adaptação da técnica *stepwise*, verificando quais componentes foram estatisticamente significativos em cada uma das configurações de ARFV. Outro importante quesito foi a diferente exposição (i.e., quantidade de elementos) em cada ARFV: como as distribuições de Poisson e Binomial Negativa demandam variável dependente discreta, não foi possível aplicar diretamente a variável número de acidentes dividido pela quantidade de elementos, assim, incorporou-se o número de elementos como uma variável de exposição (incluída como *offset* no modelo). O modelo estima, então, a frequência de acidentes esperada em um intervalo de 15 minutos a partir das dadas variáveis explicativas. Assim como as análises apresentadas até o momento, a modelagem também foi realizada segregando os acidentes por severidade. A variável *fiscalização eletrônica* não foi estatisticamente significativa em nenhum dos modelos, provavelmente por estar presente em apenas 5% dos RFV.

##### **4.4.1.1 Modelagem por ARFV: acidentes com danos materiais**

Conforme as configurações de ARFV definidas, a Tabela 11 traz as variáveis que foram significativas a um nível de 5% ao aplicar a modelagem com a técnica *stepwise*. No modelo 3, por exemplo, foram incluídas as variáveis *fluxo por faixa*, *velocidade* e *número de faixas*, das quais apenas a *velocidade* não foi selecionada na composição do melhor modelo para a configuração. Apresentam-se, também, os indicadores AIC,  $2 \log\text{-likelihood}$  e  $\sigma_d$  para as distribuições de Poisson e Binomial Negativa.

Em termos gerais, verifica-se que os modelos Binomiais Negativos apresentaram melhores indicadores do que os de Poisson (i.e., menores AIC, maiores  $2 \log\text{-likelihood}$  e  $\sigma_d$  mais próximos a um). Com relação às configurações testadas, os modelos 1 a 6 contemplam o *fluxo por faixa*, os modelos 7 a 9 incorporam o *grau de saturação* e o modelo 10 confronta as duas variáveis visando obter qual está explicando o fenômeno com mais eficácia. As variáveis *número de faixas* e *ciclos por hora* foram selecionadas pelo método e possuíram significância estatística em todos os modelos em que estavam presentes. O mesmo ocorreu para o *fluxo por*

*faixa* e para o *grau de saturação*, exceto no último modelo que sugeriu a incorporação do *fluxo por faixa*.

Tabela 11 – ARFV – acidentes com danos materiais: Variáveis significativas por configuração aplicando modelagem de Poisson e Binomial Negativa

Variáveis e indicadores	Modelo										
	1	2	3	4	5	6	7	8	9	10	
<i>#Observações</i>	32	32	32	128	128	128	160	20	80	160	
<i>Fluxo por faixa</i>	*	*	*	*	*	*				*	
<i>Velocidade</i>	-	-	-	-	-	-	-				
<i>nfx</i>	*			*		*	*	*	*	*	
<i>Ciclos por hora</i>		*		*	*				*		
<i>Relação prin-sec</i>			-		-	-					
<i>Grau de saturação</i>							*	*	*	-	
Poisson	AIC	146	166	176	342	417	373	323	97	239	319
	2 log-likelihood	-138	-156	-172	-328	-407	-365	-309	-83	-219	-311
	$\sigma_d$	1,87	1,40	1,34	1,25	1,10	1,30	1,05	1,86	1,20	1,04
Binomial Negativa	AIC	144	168	177	341	418	372	317	99	241	314
	2 log-likelihood	-134	-156	-171	-325	-406	-362	-301	-83	-219	-304
	$\sigma_d$	1,10	1,40	1,03	1,07	1,00	1,06	0,81	1,86	1,18	0,82

(\*)Variáveis selecionadas pelo *stepwise* (AIC) e com significância estatística < 5%. (-) Variáveis incluídas nos modelos, mas não selecionadas pelo método

Fonte: Elaborada pelo autor

Como indicaram as análises exploratórias efetivadas, a *velocidade* não mostrou importância para o entendimento dos acidentes com danos materiais. A ocorrência desse tipo de acidente está mais relacionada à exposição veicular do que propriamente à *velocidade* desenvolvida pelos condutores nos ambientes com alta densidade de semáforos. Hipóteses sobre a associação da *relação fluxo principal-secundária* aos acidentes podem ser levantadas em duas frentes: (i) vias principais e secundárias com fluxos veiculares similares pode aumentar a probabilidade de ocorrência de colisões – principalmente do tipo transversal – por incrementar a exposição ao conflito na área do cruzamento; (ii) vias principais com fluxos veiculares significativamente superiores aos da secundária podem estar relacionadas a aumentos nos avanços semafóricos, incrementando o risco na interseção. Entretanto a *relação*

*fluxo principal-secundária* não foi significativa em nenhum dos modelos testados. A Tabela 12 traz os coeficientes dos modelos com as variáveis selecionadas.

Tabela 12 – ARFV – acidentes com danos materiais: coeficientes dos modelos de distribuição

Binomial Negativa

Variáveis e indicadores	Modelo		
	1	2	4
<i>#Observações</i>	32	32	121
<i>Intercepto</i>	<b>-1,02E+01</b> 1,85E-01	<b>-9,43E+00</b> 2,11E-01	<b>-9,65E+00</b> 2,49E-01
<i>Fluxo por faixa</i>	<b>2,57E-03</b> 4,71E-04	<b>1,65E-03</b> 4,72E-04	<b>1,79E-03</b> 5,19E-04
<i>Ciclo por hora (30 a 37)</i>	- -	<b>-2,77E-01</b> 1,19E-01	<b>-2,85E-01</b> 1,37E-01
<i>Ciclo por hora (40 a 50)</i>	- -	<b>-4,02E-01</b> 2,20E-01	<b>-4,48E-01</b> 1,83E-01
<i>Ciclo por hora (56 a 90)</i>	- -	<b>-4,90E-01</b> 1,56E-01	<b>-4,14E-01</b> 1,96E-01
<i>nfx (5 ou 6)</i>	1,83E-01 1,44E-01	- -	1,47E-01 1,22E-01
<i>nfx (7 ou 8)</i>	3,38E-01 2,35E-01	- -	3,26E-01 2,22E-01
<i>nfx (10)</i>	<b>1,71E+00</b> 2,15E-01	- -	<b>1,69E+00</b> 1,99E-01
AIC	161,9	167,5	341,0
2 log-likelihood	-149,9	-155,5	-325,0
$\sigma_a$	1,01	1,40	1,07
EMA	3,40	2,68	1,72

<sup>1</sup>Coefficiente, <sup>2</sup>Erro padrão do coeficiente

Negrito: valor-p < 0,05; itálico: valor-p < 0,10

Fonte: Elaborada pelo autor

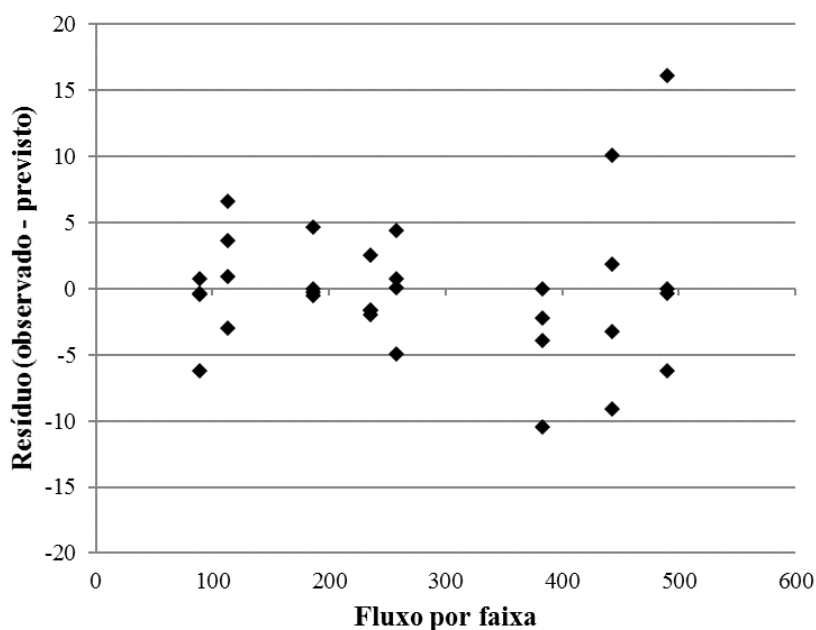
A comparação entre os indicadores de erro médio absoluto (EMA) dos três modelos mostra que o modelo 4 foi o que apresentou menores resíduos entre o previsto e o observado. Entretanto a redução no EMA com o aumento do número de observações da configuração já era esperada. A utilização de 121 configurações de ARFV, por exemplo, fez com que alguns desses agrupamentos tenham poucos elementos e, provavelmente, poucos acidentes relacionados a esses. Então, como a modelagem foi baseada na exposição, é provável que parcela dessas 121 configurações apresentem baixos valores de acidentes observados e

previstos, diluindo os resíduos daquelas configurações com maior número de elementos e de acidentes previstos e observados.

Foi mostrado na seção anterior (Figura 35) que existe uma correlação entre *fluxo por faixa* e *ciclos por hora* de -0,65, indicando correlação média-alta entre as variáveis; fato que alterou os valores dos coeficiente. A inclusão do *ciclos por hora* no modelo 4 proporcionou uma redução de 30% no coeficiente do *fluxo por faixa* quando comparado ao modelo 1, sendo um indicativo de que existe um fator comum entre as duas variáveis que está explicando a mesma parcela do efeito, impossibilitando a interpretação da influência real de cada variável. Assim, como o principal intuito desta etapa de modelagem não é a obtenção de um modelo para realizar previsões e sim uma análise confirmatória das hipóteses levantadas nas seções anteriores, decidiu-se por investigar mais detalhadamente o modelo 1.

Os 432 acidentes com danos materiais utilizados na modelagem divididos pelos 32 ARFV resultaram em uma média de 13,5 acidentes por agrupamento que comparada ao EMA de 3,4 mostra que o modelo, de certa forma, está alinhado aos dados observados de acidentes. Entretanto essa é uma avaliação global dos resíduos, sendo a investigação localizada dos erros apresentada na Figura 39 que contém o gráfico de resíduos por observação do modelo.

Figura 39 – ARFV – acidentes com danos materiais: resíduos do modelo 1



Fonte: Elaborada pelo autor

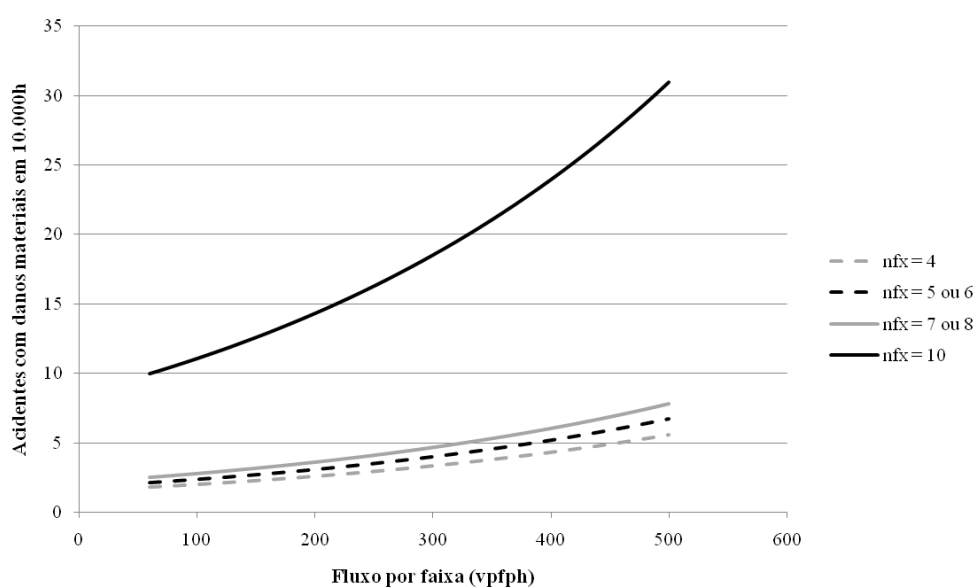
Os resíduos foram calculados com base na diferença entre os acidentes observados e os previstos pelo modelo, ou seja, valores de resíduos negativos indicam que o modelo está



superestimando a frequência de acidentes. A análise do gráfico mostra que na faixa de *fluxo por faixa* de 80 a 300 vphpf os resíduos estão bem distribuídos em valores positivos e negativos e não apresentam notáveis amplitudes. Já a categoria de 400 a 500 vphpf apresenta maiores amplitudes dos resíduos, indicando que em algumas observações há consideráveis erros no modelo.

Os coeficientes do modelo 1 confirmam que o aumento do *fluxo por faixa* está relacionado a maiores frequências de acidentes com danos materiais. Como o *número de faixas* foi incluído como fatores no modelo, os coeficientes das categorias de 5 ou 6 faixas, 7 ou 8 faixas e 10 faixas são estimados a partir do modelo “base” calibrado para a categoria de 4 faixas. As categorias 5 ou 6 faixas e 7 ou 8 faixas não mostraram significância estatística em comparação à categoria base, enquanto o segmento com 10 faixas foi significativo. A Figura 40 traz as curvas de previsão do modelo.

Figura 40 – ARFV – acidentes com danos materiais: representação gráfica do modelo 1



Fonte: Elaborada pelo autor

A Figura 40 mostra que a ordem de grandeza da categoria de 10 faixas é significativamente superior aos outros segmentos. Por exemplo, para um *fluxo por hora* de 300 vphpf estima-se cerca de quatro acidentes a cada 10.000 horas para RFV com 4 a 8 faixas, enquanto a estimativa para regimes com 10 faixas é de 17. Entretanto, destaca-se que apenas 81.079 RFV – aproximadamente 2% da amostra completa – foram utilizados para calibrar a categoria com 10 faixas. A ausência de significância estatística para as duas categorias de *número de faixas* é notada no gráfico com a proximidade entre as curvas.

Como citado na seção 4.3.2, caso o modelo selecionado possuísse poucas variáveis explicativas, permitindo uma maior desagregação da variável de exposição *fluxo por faixa*, a configuração com 17 ARFV de *fluxo por faixa* e *velocidade* também seria testada na modelagem. A aplicação do modelo gerou coeficiente para o *fluxo por faixa* de 2,66E-03; valor 3,5% superior ao coeficiente encontrado para  $K = 8$ . Os coeficientes do *número de faixa* praticamente não sofreram alterações. Assim, considera-se que com a configuração utilizada de oito ARFV foi capaz de captar satisfatoriamente a influência do *fluxo por faixa*.

#### 4.4.1.2 Modelagem por ARFV: acidentes com vítimas

Seguindo o mesmo procedimento adotado para os acidentes com danos materiais, a Tabela 13 traz as variáveis significativas por configuração de ARFV. Ao realizar a calibração utilizando a distribuição Binomial Negativa, o parâmetro de superdispersão ( $\phi$ ) não convergiu em nenhum dos modelos, indicando que a dispersão considerada na distribuição de Poisson foi suficiente para a modelagem, fato que é corroborado com os valores de  $\sigma_d$  obtidos.

Tabela 13 – ARFV – acidentes com vítimas: Variáveis significativas por configuração aplicando modelagem de Poisson

Variáveis e indicadores	Modelo									
	1	2	3	4	5	6	7	8	9	10
#Observações	32	32	32	95	128	128	160	159	159	16
Fluxo por faixa	-	-	-	-	-	-				
Velocidade	-	-	-	-	-	-	-	-	-	
<i>nfx</i>	*			*		*	*			*
Ciclos por hora		*		*	*			*		*
Relação prin-sec			-		-	-			-	
Grau de saturação							-	-	*	
AIC	117.6	143.3	149.3	278.5	346.0	292.4	264.0	313.9	335.9	72.3
2 log-likelihood	-109.6	-135.3	-145.3	-266.5	-338.0	-284.4	-258.0	-305.9	-325.9	-60.3
$\sigma_d$	1.19	1.20	0.98	1.07	0.93	0.89	0.87	0.83	0.79	0.88

\*Variáveis selecionadas pelo *stepwise* (AIC) e com significância estatística < 5%

Fonte: Elaborada pelo autor

Ao avaliar a Tabela 13, fica claro que as variáveis *ciclos por hora* e *número de faixas* foram aquelas mais relevantes para explicação do fenômeno. A análise das Figuras 34, 36 e 37 mostram que os acidentes com vítimas estão distribuídos em regimes de baixas e altas demandas veiculares, sugerindo que a oscilação nos valores de variáveis como *fluxo por faixa*, *velocidade* e *grau de saturação* não possuem fortes tendências de incremento ou redução da frequência de acidentes; fato que é confirmado com a ausência de significância estatística dessas variáveis nos modelos calibrados. Novamente surge a discussão apresentada nas seções anteriores de que uma parcela dos acidentes com vítimas é decorrente do desempenho de altas velocidades em condições de baixa demanda veicular e outra parcela ocorre em regimes congestionados, que não possibilitam o desenvolvimento de velocidade, afetando, principalmente, os usuários vulneráveis do sistema de transportes. A Tabela 14 traz os coeficientes do modelo 10.

Tabela 14 – ARFV – acidentes com vítimas: coeficientes do modelo de distribuição de

Poisson	
Variáveis e indicadores	Modelo
	10
<i>#Observações</i>	16
<i>Intercepto</i>	<b>-9.76E+00<sup>1</sup></b> 1.53E-01 <sup>2</sup>
<i>Ciclo por hora (30 a 37)</i>	-2.89E-01 1.60E-01
<i>Ciclo por hora (40 a 50)</i>	<b>-5.75E-01</b> 2.07E-01
<i>Ciclo por hora (56 a 90)</i>	<b>-4.08E-01</b> 1.85E-01
<i>nfx (5 ou 6)</i>	2.89E-01 1.51E-01
<i>nfx (7 ou 8)</i>	<b>7.16E-01</b> 2.60E-01
<i>nfx (10)</i>	<b>1.63E+00</b> 2.66E-01
AIC	90.9
2 log-likelihood	-76.9
$\sigma_a$	1.33
EMA	2.18

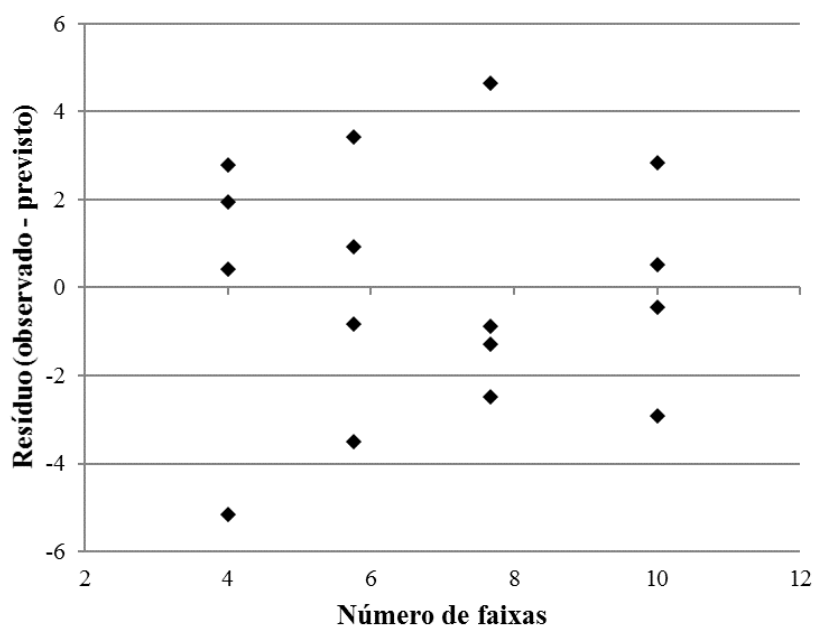
<sup>1</sup>Coefficiente, <sup>2</sup>Erro padrão do coeficiente

Negrito: valor-p < 0,05; itálico: valor-p < 0,10

Fonte: Elaborada pelo autor

Os 236 acidentes com vítimas utilizados na calibração do modelo divididos pelos 16 ARFV resultaram em um valor médio de 14,7 acidentes por configuração, assim o EMA obtido de 2,2 indica que – globalmente – o modelo está desenvolvendo previsões satisfatórias. Em comparação ao modelo de acidentes com danos materiais, a relação valor médio de acidentes por configuração e EMA aponta melhores ajustes para o modelo com vítimas (13,5/3,4 para o primeiro e 14,7/2,2 para o segundo). Esse achado é, de certa forma, surpreendente, visto que com o maior tamanho amostral dos acidentes com danos materiais – reduzindo a aleatoriedade do fenômeno – e com a mais clara relação entre as variáveis explicativas e os acidentes sem vítimas durante a análise exploratória, imaginava-se que a modelagem dos acidentes com danos materiais fosse apresentar menores resíduos em comparação à dos acidentes com vítimas. Em termos de resíduos locais (Figura 41), verifica-se que com o incremento no *número de faixas* não se percebe alterações significativas na ordem de grandeza dos erros.

Figura 41 – ARFV – acidentes com vítimas: resíduos do modelo selecionado



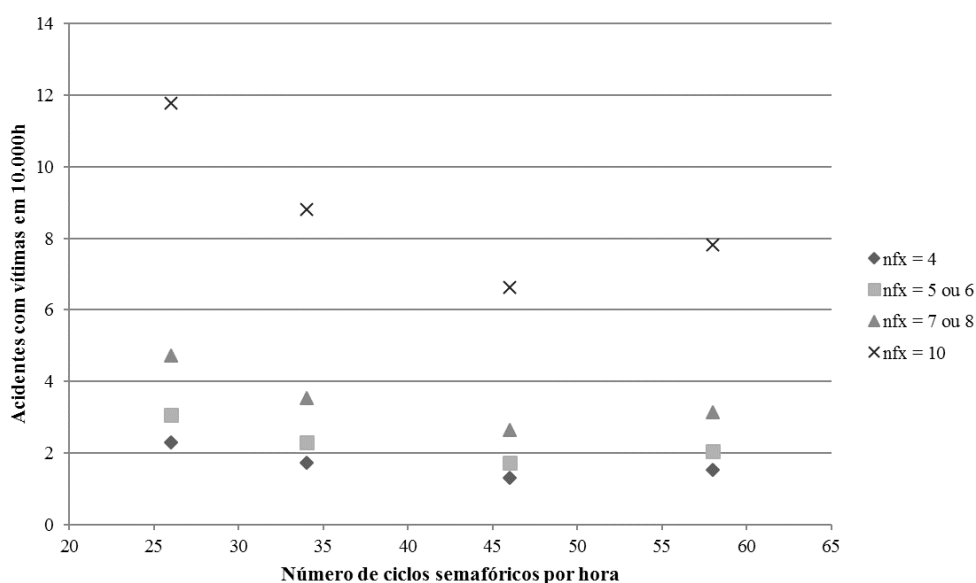
Fonte: Elaborada pelo autor

A Figura 42 traz a representação gráfica do modelo 10. Em termos de *número de faixas*, como mostrado nas seções anteriores, o aumento na variável está relacionado a maiores frequências de acidentes com vítimas, fato que pode ser atribuído às seguintes consequências do aumento no *número de faixas*: (i) incremento no número de corredores virtuais para motociclistas, podendo aumentar a interação motociclistas-veículos; (ii)

incremento de possibilidades de movimentação lateral por parte dos veículos, sendo capaz de aumentar os conflitos laterais; (iii) aumento da distância de travessia para pedestres, incrementando a exposição aos atropelamentos.

Com relação à variável *ciclos por hora*, os resultados dos modelos confirmam parcialmente as hipóteses levantadas a partir da Figura 37, a qual sugere que as maiores frequências de acidentes com vítimas ocorrem nos RFV com ciclos semafóricos mais curtos (22 a 28 *ciclos por hora*) e mais longos (56 a 90 *ciclos por hora*). O modelo confirma que as maiores frequências de acidentes estão nos regimes de ciclos semafóricos mais curtos, mas verifica-se que a estimativa para a classe de RFV mais longos é ligeiramente inferior quando comparada a do segmento de 30 a 37 *ciclos por hora*. Destaca-se, ainda, que apesar de não ter sido variável significativa em nenhuma das configurações, o *fluxo por faixa* apresenta média-alta correlação com o *ciclos por hora*, podendo uma parcela dos coeficientes atribuídos à variável corresponderem a características da exposição e não da duração do ciclo.

Figura 42 – ARFV – acidentes com vítimas: representação gráfica do modelo 10



Fonte: Elaborada pelo autor

#### 4.4.2 Modelagem estatística: abordagem por entidades

A modelagem estatística por entidades seguiu uma abordagem tradicional em que cada observação representa uma interseção semaforizada, associando as variáveis agregadas em valores médios diários ao somatório de acidentes ocorridos entre os anos de 2015 e 2017, ou seja, a previsão dos modelos resulta na estimativa da frequência de acidentes em três anos.

Como as diferentes configurações de modelos possuem o mesmo número de observações, nesta modelagem os indicadores AIC e  $2 \log\text{-likelihood}$  podem ser comparados. A técnica *forward* foi aplicada para seleção das principais variáveis e dos modelos mais robustos.

Destaca-se que, usualmente, a variável de exposição é aplicada como Volume Diário Médio Anual (VDMA), entretanto – para compatibilizar com os resultados obtidos para os ARFV – o volume veicular foi tratado como Fluxo por Faixa Médio Anual (FFMA), que representa o VDMA dividido por 24 horas.

A Figura 43 traz a matriz de correlação entre as variáveis e os acidentes por entidade. Com relação aos acidentes com danos materiais e com vítimas, as variáveis que apresentaram maiores correlações lineares foram *FFMA*, *grau de saturação*, *paradas por faixa*, *ciclos por hora*, *número de faixas*, *número de aproximações* e *número de conversões*. Assim como na análise por RFV, os grupos de variáveis (i) *FFMA*, *grau de saturação* e *paradas por faixa* e (ii) *nfx*, *app* e *conversões* exibiram alta correlação entre si, indicativo de que as mesmas não devem ser incorporadas aos modelos simultaneamente. Do mesmo modo que a análise anterior, a *fiscalização eletrônica* – possivelmente por estar presente em apenas quatro das 94 interseções – não mostrou significância estatística nos modelos.

Figura 43 – Entidade-anual: matriz de correlação

Variáveis	Danos Materiais	Vítimas	FFMA	Velocidade	Grau de saturação	Paradas por faixa	Rel. princ-sec	Ciclos por hora	Nfx	App	Conversões	FE
Danos Materiais	-											
Vítimas	0.63	-										
FFMA	0.27	0.27	-									
Velocidade	-0.07	-0.10	-0.22	-								
Grau de saturação	0.38	0.47	<b>0.71</b>	-0.31	-							
Paradas por faixa	0.28	0.35	<b>0.88</b>	-0.47	<b>0.81</b>	-						
Rel. princ-sec	0.07	0.10	-0.07	0.18	0.15	-0.15	-					
Ciclos por hora	-0.26	-0.27	-0.56	0.24	-0.41	-0.43	-0.17	-				
Nfx	0.62	0.49	0.11	0.00	0.31	0.13	0.30	-0.12	-			
App	0.48	0.40	0.10	-0.04	0.28	0.13	0.30	-0.08	<b>0.93</b>	-		
Conversões	0.52	0.47	0.16	-0.03	0.25	0.20	0.08	-0.15	<b>0.81</b>	<b>0.83</b>	-	
FE	0.14	0.02	0.08	0.08	0.06	0.02	-0.11	-0.05	0.13	0.08	0.02	-

Fonte: Elaborada pelo autor

#### 4.4.2.2 Modelagem por entidades: acidentes com danos materiais

A calibração dos modelos foi realizada utilizando uma amostra de 894 acidentes com danos materiais, representando uma média de 9,5 acidentes por entidade. Os indicadores  $\sigma_d$  apresentaram valores mais próximos a um com a distribuição Binomial Negativa, sugerindo melhor adequação dessa distribuição em comparação à de Poisson, a qual mostrou valores de  $\sigma_d$  na ordem de 2,1. A Tabela 15 traz os coeficientes e indicadores de cada modelo aplicando a técnica *forward*. Observa-se que, diferentemente da modelagem anterior, nenhuma das variáveis foi inserida como fator.

Tabela 15 – Entidades – acidentes com danos materiais: coeficientes e indicadores dos modelos com distribuição Binomial Negativa

Variáveis e indicadores	Modelo					
	1	2	3	4	5	6
Intercepto	<b>1.36E+00<sup>1</sup></b>	<b>1.40E+00</b>	3.78E-01	<b>2.42E+00</b>	<b>1.22E+00</b>	<b>8.25E-01</b>
	2.90E-01 <sup>2</sup>	4.22E-01	2.89E-01	6.09E-01	3.18E-01	2.02E-01
Fluxo por faixa médio anual - FFMA	<b>3.85E-03</b>	<b>3.81E-03</b>	<b>3.52E-03</b>	2.22E-03	<b>3.97E-03</b>	-
	1.23E-03	1.26E-03	1.02E-03	1.47E-03	1.23E-03	-
Velocidade	-	-1.57E-03	-	-	-	-
	-	1.43E-02	-	-	-	-
Número de faixas	-	-	<b>1.88E-01</b>	-	-	<b>1.62E-01</b>
	-	-	2.75E-02	-	-	2.97E-02
Ciclos por hora	-	-	-	-1.77E-02	-	-
	-	-	-	9.09E-03	-	-
Relação fluxo principal-secundária	-	-	-	-	4.70E-02	-
	-	-	-	-	4.73E-01	-
Grau de saturação	-	-	-	-	-	<b>1.12E-02</b>
	-	-	-	-	-	3.76E-03
AIC	570,4	572,4	534,6	568,8	571,4	537,6
2 log-likelihood	-564,4	-564,4	-526,6	-560,8	-563,4	-529,6
$\sigma_d$	1,02	1,03	1,03	1,02	1,03	1,02
EMA	4,19	4,48	3,48	4,18	4,13	3,70

<sup>1</sup>Coefficiente, <sup>2</sup>Desvio padrão do coeficiente. Negrito: valor-p < 0,05

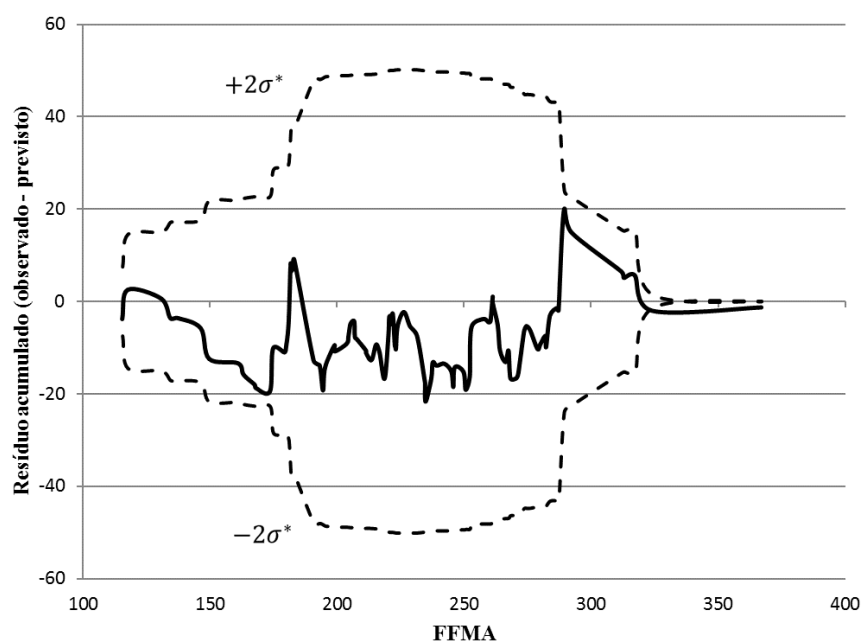
Fonte: Elaborada pelo autor

Apesar de não ser a variável de maior correlação com os acidentes com danos materiais, devido a sua importância e representatividade nos modelos observados na revisão da literatura, decidiu-se por utilizar o *FFMA* como variável base do processo (modelo 1). A inclusão sequencial das outras variáveis mostrou que apenas o *número de faixas* (modelo 3) trouxe melhorias significativas ao modelo base. Como o *grau de saturação* apresentou

considerável correlação com a variável dependente, considerou-se essa variável no modelo 6, junto do *número de faixas*. Em termos gerais, percebe-se que os modelos 3 e 6 foram os que exibiram melhores indicadores, com leve vantagem para o três.

A interpretação dos coeficientes mostra que tanto o aumento do *FFMA* quanto do *número de faixas* incrementa o número previsto de acidentes. Maior detalhamento dessa relação é apresentado na seção seguinte, quando os resultados de cada modelagem são confrontados. Em termos de resíduos locais, a Figura 44 traz o *CuRe plot* do modelo 3. De modo geral, a linha de resíduos acumulados não rompe as envoltórias do desvio, entretanto percebe-se que em quase toda a extensão do gráfico os resíduos acumulados estão abaixo do eixo das abscissas. Isso ocorre devido à superestimação das previsões no intervalo de *FFMA* entre 130 e 170 vphpf. Nas outras faixas de *FFMA*, os resíduos tendem a oscilar em torno de um mesmo valor, indicando boa adequação do modelo aos dados observados.

Figura 44 – Entidades – acidentes com danos materiais: *CuRe plot* modelo 3



Fonte: Elaborada pelo autor

#### 4.4.2.2 Modelagem por entidades: acidentes com vítimas

A calibração dos modelos foi realizada utilizando uma amostra de 499 acidentes com vítimas, representando uma média de 5,3 acidentes por entidade. Novamente, os indicadores  $\sigma_d$  apresentaram valores mais próximos a um com a distribuição Binomial Negativa. O  $\sigma_d$



considerando a distribuição de Poisson girou em torno de 1,6. A Tabela 16 traz os coeficientes e indicadores de cada modelo aplicando a técnica *forward*.

Tabela 16 – Entidades – acidentes com vítimas: coeficientes e indicadores dos modelos com distribuição Binomial Negativa

Variáveis e indicadores	Modelo							
	1	2	3	4	5	6	7	8
Intercepto	<b>8.8E-01<sup>1</sup></b>	<b>1.0E+00</b>	-5.9E-03	<b>2.2E+00</b>	<b>7.1E-01</b>	2.6E-01	<b>1.2E+00</b>	<b>1.7E+00</b>
	3.3E-01 <sup>2</sup>	4.8E-01	3.5E-01	6.9E-01	3.6E-01	2.3E-01	4.6E-01	3.7E-01
Fluxo por faixa médio anual - FFMA	<b>3.4E-03</b>	<b>3.3E-03</b>	<b>3.3E-03</b>	1.4E-03	<b>3.5E-03</b>	-	-	-
	1.4E-03	1.4E-03	1.3E-03	1.7E-03	1.4E-03	-	-	-
Velocidade	-	-7.6E-03	-	-	-	-	-	-
	-	1.6E-02	-	-	-	-	-	-
Número de faixas	-	-	<b>1.6E-01</b>	-	-	<b>1.1E-01</b>	-	<b>1.6E-01</b>
	-	-	3.3E-02	-	-	3.3E-02	-	3.3E-02
Ciclos por hora	-	-	-	<b>-2.1E-02</b>	-	-	-1.2E-02	<b>-2.5E-02</b>
	-	-	-	1.0E-02	-	-	8.7E-03	8.1E-03
Relação fluxo principal-secundária	-	-	-	-	6.4E-02	-	-	-
	-	-	-	-	5.3E-02	-	-	-
Grau de saturação	-	-	-	-	-	<b>1.7E-02</b>	<b>1.9E-02</b>	-
	-	-	-	-	-	4.2E-03	4.6E-03	-
AIC	484	485	465	482	484	458	467	462
2 log-likelihood	-478	-477	-457	-474	-476	-450	-459	-454
$\sigma_d$	1,07	1,08	1,10	1,08	1,08	1,11	1,08	1,10
EMA	2,59	2,62	2,32	2,57	2,59	2,23	2,33	2,34

<sup>1</sup>Coefficiente, <sup>2</sup>Desvio padrão do coeficiente. Negrito: valor-p < 0,05

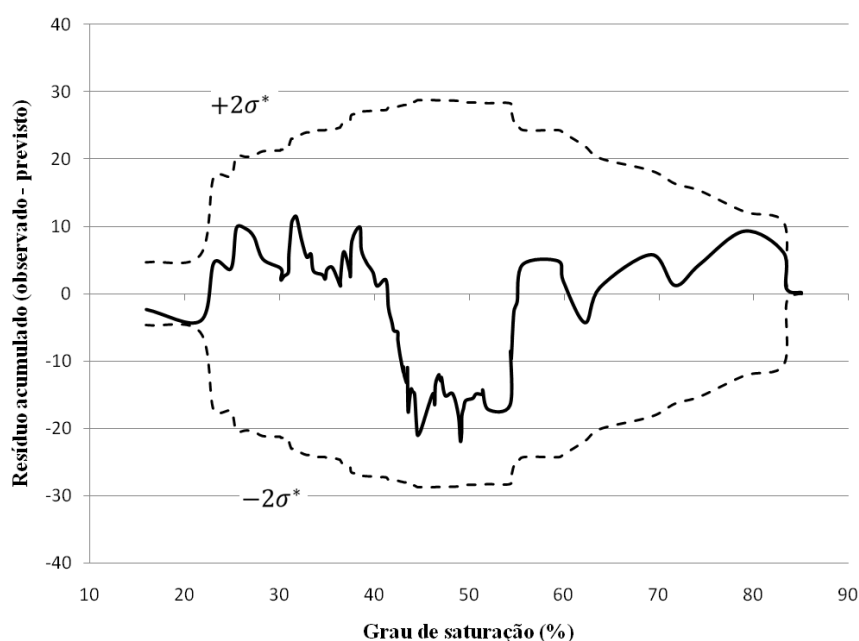
Fonte: Elaborada pelo autor

Seguindo o procedimento adotado para a modelagem dos acidentes com danos materiais, aplicou-se o modelo base com a variável *FFMA*, sucedendo com a inclusão sequencial das variáveis e com a utilização de configurações com base no *grau de saturação*. Novamente, o modelo 3 com o *FFMA* que apresentou melhores resultados foi o que contém o *número de faixas*. Percebe-se, também, que a inclusão do *ciclos por hora* tornou o *FFMA* não significativo, provavelmente devido à correlação entre as duas variáveis. O modelo 6, que considera o *grau de saturação* e o *número de faixas*, exibiu melhores valores de indicadores do que o modelo 6. Por fim, considerou-se o modelo 8, cuja configuração é idêntica à selecionada na modelagem dos acidentes com vítimas para os ARFV, entretanto os indicadores do modelo 6 seguem sugerindo melhores estimativas.

A interpretação dos coeficientes mostra que tanto o aumento do *grau de saturação* quanto do *número de faixas* incrementa o número previsto de acidentes. A Figura 45 traz o

*CuRe plot* do modelo melhor avaliado. Em termos gerais, verifica-se que em nenhum momento os resíduos acumulados ultrapassam as envoltórias do desvio, indicando que os erros estão dentro de uma ordem de grandeza tolerável. Notam-se que na faixa de 41 a 43% do *grau de saturação* há uma tendência de superestimação das previsões – verificada na queda da linha dos resíduos –, e que na ordem de 55% existem algumas observações em que o número previsto de acidentes foi consideravelmente inferior ao observado.

Figura 45 – Entidades – acidentes com vítimas: *CuRe plot* modelo 6

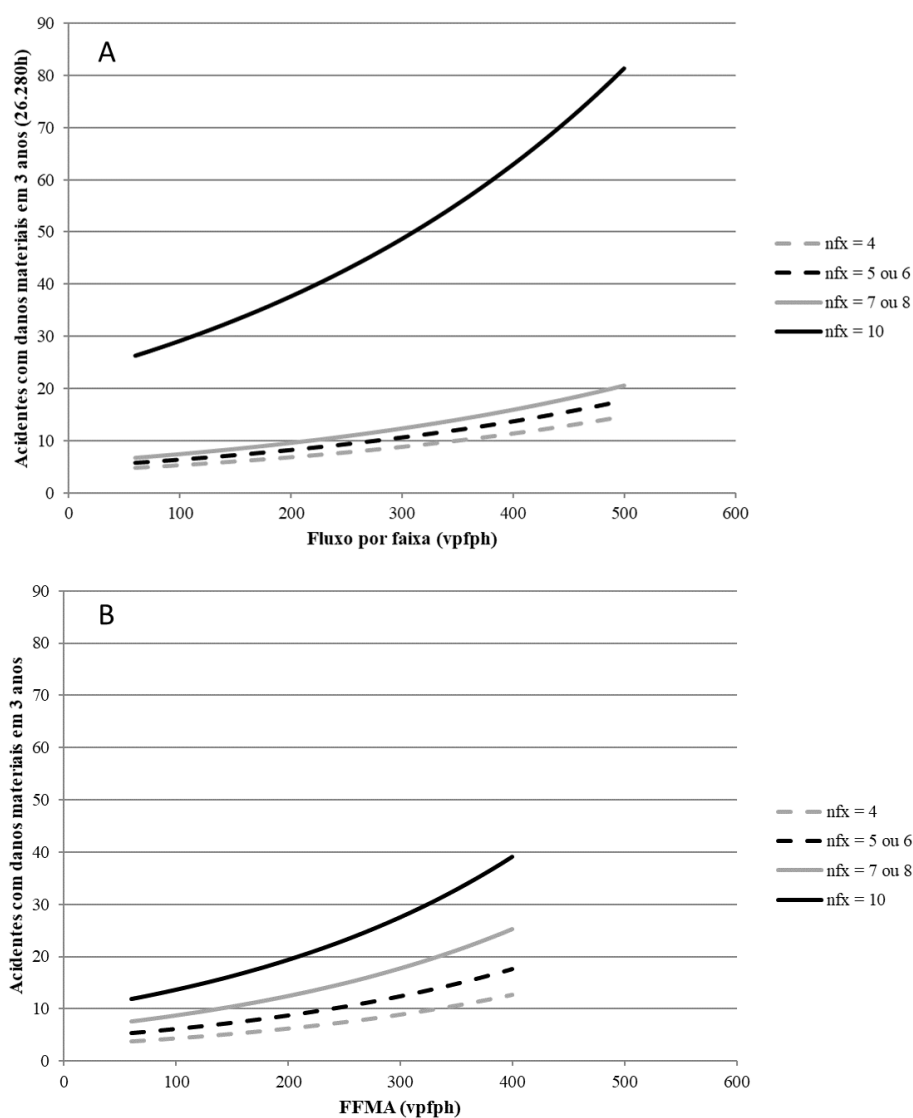


Fonte: Elaborada pelo autor

#### 4.4.3 Comparação entre as modelagens estatísticas: ARFV e entidades

A comparação entre as modelagens por ARFV e por entidades foi realizada considerando três quesitos: (i) variáveis significativas; (ii) coeficientes e estimações; (iii) interpretação dos resultados. Com relação à modelagem dos acidentes com danos materiais, as duas abordagens resultaram nas mesmas variáveis: *fluxo por faixa* e *número de faixas*. Os coeficientes de ambas as variáveis foram positivos nas duas modelagens, indicando que o incremento da variável está relacionado ao aumento na estimativa da frequência de acidentes. A Figuras 46A e 47B trazem as estimativas dos acidentes com danos materiais a partir dos modelos por ARFV e entidades, respectivamente, para um período de três anos (ou 26.280 horas).

Figura 46 – Acidentes com danos materiais: comparação das estimativas por abordagem. (A) ARFV e (B) entidades-anual



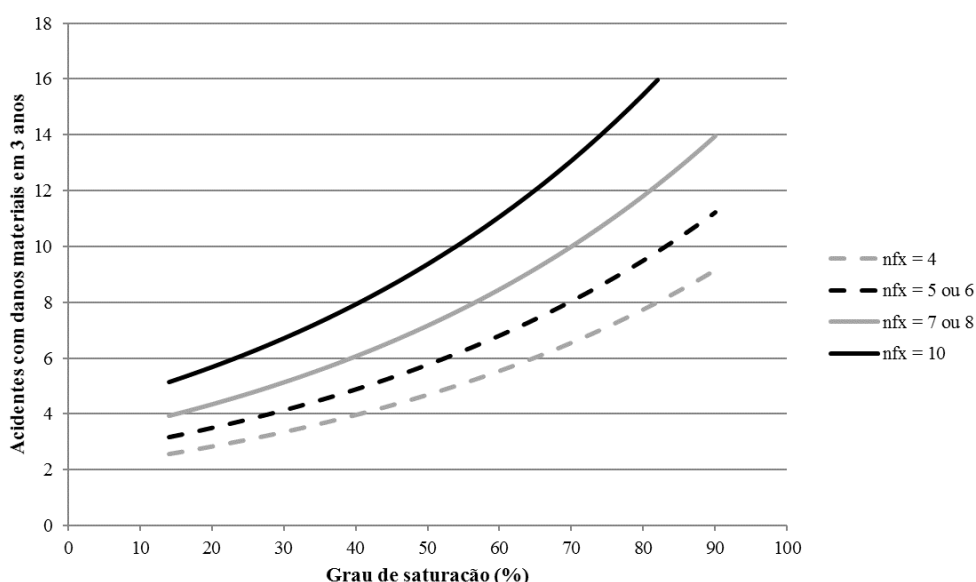
Fonte: Elaborada pelo autor

A comparação entre as estimativas por abordagem mostra que os resultados para as categorias de 4 a 8 faixas são – de certo modo – similares. Por exemplo, em RFV com *fluxo por faixa* de 300 vphpf e 5 ou 6 *faixas* estima-se que em três anos ocorrem cerca de dez acidentes com danos materiais, já em uma interseção semaforizada com *FFMA* de 300 vphpf e 5 ou 6 *faixas* a estimativa para o mesmo intervalo é de 12 acidentes. Ao analisar a categoria de 10 faixas percebe-se que a frequência estimada para a modelagem por ARFV é consideravelmente superior à por entidades, principalmente para as faixas mais elevadas de *fluxo por faixa*.

Pode-se interpretar que, de forma desagregada, nos RFV com maior demanda veicular existe uma maior exposição aos acidentes tornando as ocorrências com danos materiais mais prováveis e que, de forma agregada, aquelas interseções semaforizadas que apresentaram os maiores *VDMA* foram as que também concentraram as mais críticas frequências de acidentes com danos materiais.

A modelagem dos acidentes com vítimas, por sua vez, trouxe diferentes variáveis significativas por tipo de abordagem: enquanto o tratamento por ARFV resultou no *ciclos por hora* e no *número de faixas* como principais variáveis, a abordagem por entidades trouxe que o *grau de saturação* e o *número de faixas* foram os principais componentes do modelo. A Figura 47 traz as previsões para o modelo por entidades.

Figura 47 – Entidades – acidentes com vítimas: representação gráfica do modelo 6

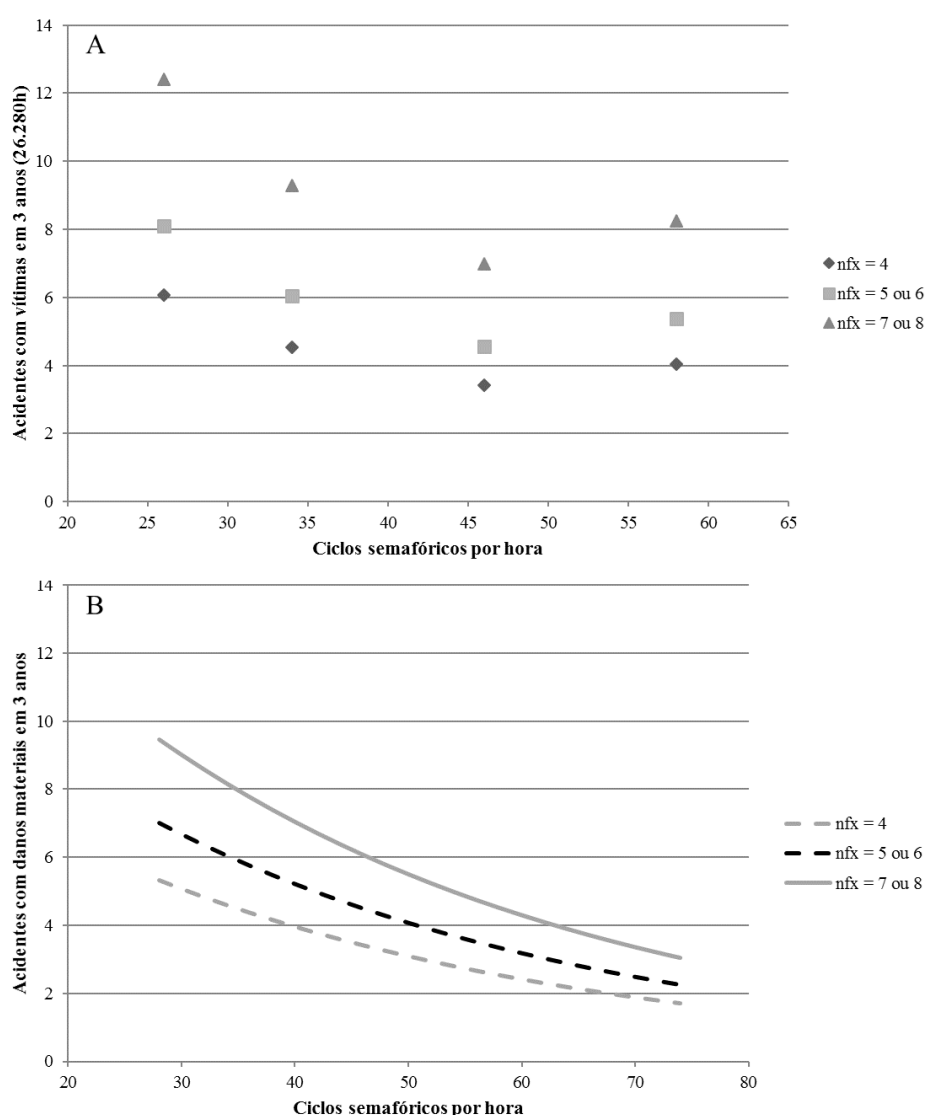


Fonte: Elaborada pelo autor

A Figura 47 mostra que, pela abordagem por entidades, os incrementos no *grau de saturação* e no *número de faixas* estão associados a maiores frequências de acidentes com vítimas. Entretanto essa associação provavelmente não está refletindo o que ocorre nas interseções de Fortaleza. Foi verificado que os acidentes com vítimas tendem a acontecer uma parte devido às altas demandas veiculares e outra graças ao desempenho de altas velocidades, partes que refletem dois diferentes aspectos de uma interseção semaforizada durante um mesmo dia. Assim, espera-se que uma variável que representa a agregação média das condições de tráfego em um local não é capaz de reproduzir estimativas confiáveis acerca da frequência de acidentes com vítimas.

Visando comparar as duas abordagens, utiliza-se do modelo 8 (Tabela 16) que apresentou indicadores similares ao melhor modelo e que faz uso das mesmas variáveis adotadas na modelagem por ARFV. As Figuras 48A e 48B trazem as estimativas dos modelos para um período de três anos a partir das abordagens por ARFV e entidades, respectivamente. Para facilitar a visualização da ordem de grandeza da escala do eixo das ordenadas, retirou-se a classe de 10 faixas de tráfego dos gráficos.

Figura 48 – Acidentes com vítimas: comparação das estimativas por abordagem. (A) ARFV e (B) entidades



Fonte: Elaborada pelo autor

A comparação entre os resultados mostra duas principais diferenças entre as abordagens: (i) na abordagem por ARFV as estimativas aumentaram na classe de maiores

*ciclos por hora* (entre 55 e 60 ciclos), enquanto na modelagem por entidades as previsões sempre diminuiriam com o aumento da variável; (ii) a ordem de grandeza das estimativas por ARFV foram maiores – principalmente das classes das extremidades – do que as previsões por entidades.

A comparação entre os resultados obtidos por cada abordagem não é simples, visto que cada uma delas produz resultados provenientes de diferentes tipos de dados. Enquanto o principal foco dos ARFV está nas características antecessoras a cada acidente, a técnica por entidades associa atributos agregados temporalmente em escala anual ao somatório da frequência de acidentes, inviabilizando que um maior detalhamento sobre cada ocorrência fosse obtido.

Além disso, a aplicabilidade de cada uma das abordagens também é diferente. A metodologia por ARFV permite o monitoramento em tempo real das probabilidades de ocorrência de acidentes, por exemplo: a incorporação dos resultados obtidos tanto na análise exploratória quanto na modelagem à tela de monitoramento do sistema *SCOOT* proporciona que condições críticas sejam identificadas na malha viária, possibilitando que medidas proativas sejam tomadas para reduzir as chances de ocorrer um acidente. Uma das condições críticas identificadas para ambas as severidades de acidentes foi a simultaneidade de *fluxos por faixa* na ordem de 400 a 500 vphpf, de *velocidades médias* na faixa de 5 a 10 km/h e de 56 a 90 *ciclos por hora*. Como a duração do ciclo semafórico pode ser controlada em tempo real pelo operador do sistema, alterações seriam recomendadas para redução do risco de acidentes. Por outro lado, os modelos baseados em entidades tradicionalmente são utilizados para comparar a frequência de acidentes observada e estimada, a partir de suas características, de entidades; possibilitando identificar os pontos críticos da malha viária. Essa abordagem tem caráter mais reativo do que a anterior por, geralmente, necessitar que os acidentes ocorram para que medidas sejam efetuadas.

## 5 CONCLUSÃO

Este trabalho apresentou uma avaliação da relação entre regimes de fluxo veicular (RFV) desagregados em intervalos de 15 minutos e a frequência e severidade de acidentes de trânsito em interseções semaforizadas da cidade de Fortaleza. As interseções estudadas estão localizadas na região central da cidade, caracterizada pela alta densidade de comércios, serviços e residências, além de ser a zona da cidade de maior adensamento de semáforos.

O primeiro objetivo específico trouxe a comparação dos horários reportados dos acidentes de trânsito e de perturbações na corrente de tráfego detectadas a partir de reduções abruptas na velocidade média estimada por equipamentos de fiscalização eletrônica, visando obter condições precursoras aos acidentes de maior confiabilidade. A aplicação do algoritmo para detecção automática das perturbações trouxe que a agregação da velocidade em três minutos aumentou consideravelmente a taxa de detecção em comparação à agregação de um minuto, pois reduziu a oscilação natural da velocidade em decorrência da presença do semáforo. Os resultados mostraram uma diferença média de 19 minutos entre os horários de reportagem dos acidentes e os horários de detecção da perturbação na velocidade. Verificou-se, também, que acidentes com vítimas apresentaram diferença média entre os horários inferior à média dos acidentes com danos materiais (valor-p = 0,01).

O segundo objetivo específico consistiu no desenvolvimento de agrupamentos homogêneos dos 4,1M RFV consolidados com informações de *fluxo por faixa, velocidade, grau de saturação, ciclos semaforicos por hora, número de faixas*, entre outros. As "configurações base" dos agrupamentos foram definidas – a partir do *fluxo por faixa* e da *velocidade* – considerando três indicadores/testes: *Silhouette test*, *Gap Statistic test* e SEQA. Os dois primeiros testes indicaram a configuração ótima de dois agrupamentos, entretanto por ir de encontro à abordagem desagregada proposta, utilizou-se do SEQA para definição do número de grupos, mesmo que esse não representasse a conformação ótima sugerida pelos outros testes. A configuração com oito agrupamentos foi selecionada para realizar as posteriores análises.

A terceira – e principal – etapa apresentou a análise exploratória e a modelagem acerca da relação entre os RFV e a frequência e severidade dos acidentes de trânsito. Os resultados obtidos possibilitaram que algumas das hipóteses levantadas na seção da revisão bibliográfica fossem corroboradas, enquanto outras não puderam ser ratificadas devido a características dos dados utilizados:

1. *RFV congestionados (altos fluxos veiculares e baixas velocidades) estão associados a maiores frequências de acidentes, principalmente os de baixa severidade*: hipótese corroborada. A análise exploratória mostrou que regimes com *fluxo por faixa* na grandeza de 400 a 600 vphpf possuem probabilidades de duas a três vezes maiores de ocorrência de acidentes com danos materiais do que condições com demandas na ordem de 100 a 300 vphpf. A modelagem trouxe que o incremento na variável *fluxo por faixa* está relacionado a aumento na frequência de acidentes prevista;
2. *RFV não congestionados (baixos-médios fluxos veiculares e médias-baixas velocidades) estão associados a menores frequências de acidentes, mas com predominância de ocorrências severas*: hipótese corroborada. A avaliação dos acidentes com vítimas apresentou maior complexidade. A análise exploratória expôs que a frequência desse tipo de acidentes é praticamente constante ao longo de todo o dia, indicando que uma parcela das ocorrências é em função da possibilidade de desempenho de velocidades em regimes de menor demanda veicular – corroborando com a hipótese – e outra parcela devido à maior exposição, principalmente dos usuários vulneráveis, em regimes de alta demanda veicular. A modelagem dos acidentes com vítimas não indicou o *fluxo por faixa* e a *velocidade* como variáveis significativas para o fenômeno;
3. *Colisões traseiras são mais frequentes em RFV de alta demanda veicular*: hipótese corroborada. A análise exploratória mostrou que há uma tendência de as colisões traseiras ocorrerem em regimes com altos valores de *fluxo por faixa* e de *grau de saturação*, valores consideravelmente superiores às condições médias dos RFV;
4. *O aumento no número de faixas está associado a maiores frequências de acidentes (CUNTO; CASTRO NETO; BARREIRA, 2011)*: hipótese parcialmente corroborada. As análises exploratórias e confirmatórias exibiram que o aumento no *número de faixas* está relacionado ao incremento nas frequências de acidentes com danos materiais e com vítimas, entretanto a utilização da variável de exposição em função da quantidade de faixas faz com que o aumento no *número de faixas* também represente um significativo incremento na quantidade total de veículos no local, aumentando também a exposição aos acidentes;
5. *O aumento no número de aproximações está relacionado a menores frequências de acidentes (XIE et al., 2013)*: os dados não permitiram a avaliação da hipótese. Devido à alta correlação entre o *número de faixas* e o *número de aproximações*,



não foi possível isolar o efeito da variável para avaliar sua influência na frequência de acidentes;

6. *O aumento no número de ciclos semaforicos por hora, associado a maiores repetições do movimento parada-arranque, está relacionado a maiores frequências de acidentes:* hipótese corroborada. A análise exploratória da influência da *quantidade de ciclos semaforicos por hora*, associada aos resultados obtidos na etapa de modelagem, mostrou que os RFV caracterizados por (i) *baixa demanda veicular + ciclos semaforicos longos* e (ii) *alta demanda veicular + ciclos semaforicos curtos* foram os que apresentaram as mais altas probabilidades de ocorrência de acidentes com e sem vítimas. Esses regimes estão relacionados, respectivamente, ao desempenho de altas velocidades e ao incremento de conflitos longitudinais, ocasionado pelo aumento no movimento parada-arranque.

A última etapa consistiu na comparação dos resultados obtidos com as modelagens por ARFV e por entidades. A modelagem dos acidentes com danos materiais resultou nas mesmas variáveis significativas para as duas abordagens: *fluxo por faixa* e *número de faixas*. Mostrando que, de forma desagregada, a ocorrência de regimes de altas demandas veiculares aumenta a probabilidade de acontecimento de um acidente com dano material e que, de forma agregada, as interseções que apresentaram os maiores *VDMA* foram as que também concentraram as mais críticas frequências de acidentes.

Com relação aos acidentes com vítimas, a modelagem por ARFV exibiu o *ciclos por hora* e o *número de faixas* como variáveis significativas, cujos coeficientes corroboraram com as hipóteses levantadas na análise exploratória. Já a abordagem por entidades ressaltou o *grau de saturação* e o *número de faixas* como atributos mais importantes. Os coeficientes do modelo sugerem que o aumento em ambas variáveis está associado ao incremento na frequência dos acidentes.

A técnica baseada em RFV agregados em 15 minutos aplicada neste trabalho permitiu que cada acidente fosse associado a características de tráfegos antecessoras a sua ocorrência, diferentemente da abordagem por entidades que utiliza informações de cada local em agregações anuais, que impossibilita um maior refinamento no entendimento da relação entre a frequência e severidade dos acidentes com trânsito com os atributos de tráfego variáveis ao longo de um mesmo dia. O foco na condição precursora do acidente aplicada na abordagem desagregada, e não na característica média da entidade, proporcionou um maior detalhamento dessa relação.

Além disso, os resultados alcançados na abordagem por RFV permitem o desenvolvimento de aplicações para o monitoramento em tempo real das probabilidades de acontecimento de acidentes nas interseções semaforizadas de Fortaleza, possibilitando que medidas proativas possam ser efetivadas pelos gestores de tráfego da cidade visando à redução das chances de ocorrência de acidentes. Por outro lado, os resultados obtidos na abordagem por entidades usualmente são utilizados na definição de pontos críticos e em estudos do tipo antes-depois, na busca da estimativa mais precisa da frequência de acidentes; viabilizando que intervenções de caráter reativo sejam efetivadas. Acredita-se, assim, que o esforço para obtenção da relação entre os acidentes de trânsito e características de tráfego em agregações desagregadas é válido e deve ser melhor explorado para outros tipos de entidades do ambiente urbano, como segmentos viários e interseções não semaforizadas.

### **5.1 Limitações e trabalhos futuros**

O desenvolvimento deste trabalho apresentou algumas limitações na etapa do método e nos resultados obtidos:

1. A velocidade estimada pelo *SCOOT* é baseada no tempo de atraso dos veículos e agregada em intervalos de 15 minutos, impossibilitando a captação de condutores que desempenharam altas velocidades;
2. Os semáforos que possuem a tecnologia *SCOOT* na cidade de Fortaleza estão localizados em regiões de alto adensamento de comércios, serviços, residências e semáforos, limitando as conclusões obtidas no estudo para localidades com cenários similares;
3. A aplicação da “correção média” nos horários reportados dos acidentes, apesar de ter sido validada na análise complementar I, pode ter trazido algumas associações errôneas entre acidentes e características precursoras;
4. A não incorporação de variáveis que podem ser importantes para explicação do fenômeno de acidentes, principalmente dos com vítimas, como: índice de agressividade dos condutores, movimentação de motociclistas entre veículos, avanços semaforicos, desrespeitos à velocidade permitida, vulnerabilidade dos usuários, utilização de dispositivos de segurança (cinto de segurança, capacete, cadeirinha de criança), ingestão de álcool e direção, entre outras.

Os estudos aplicando dados desagregados para o ambiente urbano são escassos na literatura. A agregação temporal usada neste trabalho permitiu o desenvolvimento da relação macroscópica das características de tráfego e dos acidentes. A utilização de um banco de dados que contenha informações em nível de passagem veicular possibilita o uso de dados ainda mais desagregados, visando à investigação do efeito de oscilações pontuais na corrente de tráfego (velocidade, desvio padrão da velocidade, *headway*, entre outros) na ocorrência dos acidentes no ambiente de fluxo interrompido. Outros tipos de entidades do ambiente urbano, como interseções não semaforizadas e segmentos viários, também podem ser explorados em níveis macroscópicos e microscópicos. Esse tipo de dados também permite a utilização de softwares de microsimulação de tráfego, com o intuito de avaliar a relação entre conflitos veiculares e acidentes de trânsito. Surge também a possibilidade de uso de técnicas estatísticas mais robustas e recentes para obtenção dos agrupamentos e para desenvolvimento da modelagem.

## REFERÊNCIAS

- AARTS, L.; VAN SCHAGEN, I. Driving speed and the risk of road crashes: A review. **Accident Analysis and Prevention**, v. 38, n. 2, p. 215–224, 2006.
- ABDEL-ATY, M. et al. **Relating Crash Occurrence to Freeway Loop Detectors Data, Weather Conditions and Geometric Factors**. Florida-EUA: [s. n.], 2005.
- ABDEL-ATY, M. et al. Real-time prediction of visibility related crashes. **Transportation Research Part C: Emerging Technologies**, v. 24, p. 288–298, 2012.
- ABDEL-ATY, M.; PANDE, A. Identifying crash propensity using specific traffic speed conditions. **Journal of Safety Research**, v. 36, n. 1, p. 97–108, 2005.
- ABDEL-ATY, M.; PEMMANABOINA, R. Calibrating a real-time traffic crash-prediction model using archived weather and ITS traffic data. **IEEE Transactions on Intelligent Transportation Systems**, v. 7, n. 2, p. 167–174, 2006.
- ABDEL-ATY, M.; WANG, X. Crash Estimation at Signalized Intersections Along Corridors: Analyzing Spatial Effect and Identifying Significant Factors. **Transportation Research Record**, v. 1953, n. 1953, p. 98–111, 2006.
- AGUERO-VALVERDE, J.; JOVANIS, P. Analysis of Road Crash Frequency with Spatial Models. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2061, p. 55–63, 2008.
- AGUERO-VALVERDE, J.; JOVANIS, P. Bayesian Multivariate Poisson Lognormal Models for Crash Severity Modeling and Site Ranking. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2136, p. 82–91, 2009.
- AKAIKE, H. A New Look at the Statistical Model Identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974.
- AMC. **Relatório Anual de Segurança Viária de Fortaleza 2016**. Fortaleza: [s. n.], 2017.
- AQUINO, E. A. DE. **Validação Do Modelo Mesoscópico De Tráfego Do Scoot Para O Desenvolvimento De Redes Viárias Urbanas Microssimuladas**. Dissertação (Mestrado em Engenharia de Transportes) - Departamento de Engenharia de Transportes, Universidade Federal do Ceará, Fortaleza, 2012.
- BALLINGER, G. A. Using Generalized Estimating Equations for Longitudinal Data Analysis. **Organizational Research Methods**, v. 7, n. 2, p. 127-150, 2004.
- CALEFFI, F. **Proposição de um Método de Harmonização da Velocidade Baseado em Modelo de Previsão de Conflitos Veiculares**. Tese (Doutorado em Engenharia de Sistemas de Transportes) - Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, 2018.
- CASTRO, M.; PALETI, R.; BHAT, C. R. A latent variable representation of count data

models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. **Transportation Research Part B: Methodological**, v. 46, n. 1, p. 253–272, 2012.

CEDER, A.; LIVNEH, M. Relationships between road accidents and hourly traffic flow-I. Analyses and interpretation. **Accident Analysis and Prevention**, v. 14, n. 1, p. 19–34, 1982.

CELIKOGLU, H. B. An Approach to Dynamic Classification of Traffic Flow Patterns. **Computer-Aided Civil and Infrastructure Engineering**, v. 28, n. 4, p. 273–288, 2013.

CHANG, G. L.; XIANG, H. **The relationship between congestion levels and accidents, Report MD-03-SP 208B46**. College Park-EUA: [s. n.], 2003.

CHEN, P.; SHEN, Q. Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. **Accident Analysis & Prevention**, v. 86, p. 239–246, jan. 2016.

CHEN, Y.; KIM, J.; MAHMASSANI, H. S. Pattern recognition using clustering algorithm for scenario definition in traffic simulation-based decision support systems. **17th International IEEE Conference on Intelligent Transportation Systems (ITSC)**, n. October, p. 798–803, 2014.

CHILUKURI, B. R.; PERRIN JR., J.; MARTIN, P. T. SCOOT and Incidents: Performance Evaluation in Simulated Environment. **Transportation Research Record**, n. 1867, p. 224–232, 2004.

CHIN, H. C.; QUDDUS, M. A. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. **Accident Analysis and Prevention**, v. 35, n. 2, p. 253–259, 2003.

CHRISTOFOROU, Z.; COHEN, S.; KARLAFTIS, M. G. Identifying crash type propensity using real-time traffic data on freeways. **Journal of Safety Research**, v. 42, n. 1, p. 43–50, 2011.

CHUNG, E. Classification of traffic pattern. **Proc. of the 11th World Congress on ITS**, n. May, p. 4–6, 2003.

CUNTO, F. J. C.; CASTRO NETO, M. M.; BARREIRA, D. S. Modelos de Previsão de Acidentes de Trânsito em Interseções Semaforizadas de Fortaleza. **Transportes**, v. 20, n. 2, p. 55–62, 2011.

CUNTO, F.; SOBREIRA, L. T. P.; FERREIRA, S. Assessing the Transferability of the Highway Safety Manual Predictive Method for Urban Roads in Fortaleza City, Brazil. **Road Safety and Simulation**, v. 141, n. 1, p. 04014072, 2013.

DAVIS, G. A. Possible aggregation biases in road safety research and a mechanism approach to accident modeling. **Accident Analysis and Prevention**, v. 36, n. 6, p. 1119–1127, 2004.

DONG, C. et al. Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections. **Accident Analysis and Prevention**, v. 70, p. 320–329, 2014.

EL-BASYOUNY, K.; SAYED, T. Collision prediction models using multivariate Poisson-lognormal regression. **Accident Analysis and Prevention**, v. 41, n. 4, p. 820–828, 2009.

ELVIK, R. **The Power Model of the relationship between speed and road safety: Update and new analyses. TOI report 1034.** Oslo-NOR: [s. n.], 2009.

ELVIK, R.; CHRISTENSEN, P.; AMUNDSEN, A. **Speed and road accidents: An evaluation of the Power Model. TOI report 740.** Oslo-NOR: [s. n.], 2004.

ERMAN, J.; ARLITT, M. Traffic classification using clustering algorithms. **SIGCOMM Workshop, Pisa, Italy**, p. 281–286, 2006.

GARBER, N.; SUBRAMANYAN, S. Incorporating Crash Risk in Selecting Congestion-Mitigation Strategies: Hampton Roads Area (Virginia) Case Study. **Transportation Research Record**, v. 1746, n. 1, p. 1–5, 2001.

GARBER, N.; WU, L. **Stochastic Models Relating Crash Probabilities With Geometric and Corresponding Traffic Characteristics Data. Research Report No. UVACTS-5-15-74.** Charlottesville: [s. n.], 2001.

GARCÍA-RÓDENAS, R.; LÓPEZ-GARCÍA, M. L.; SÁNCHEZ-RICO, M. T. An Approach to Dynamical Classification of Daily Traffic Patterns. **Computer-Aided Civil and Infrastructure Engineering**, v. 32, n. 3, p. 191–212, 2017.

GITELMAN, V.; DOVEH, E.; BEKHOR, S. The Relationship between Free-Flow Travel Speeds, Infrastructure Characteristics and Accidents, on Single-Carriageway Roads. **Transportation Research Procedia**, v. 25, p. 2031–2048, 2017.

GOLOB, T. F.; RECKER, W.; PAVLIS, Y. Probabilistic models of freeway safety performance using traffic flow data as predictors. **Safety Science**, v. 46, n. 9, p. 1306–1333, 2008.

GOMES, M. J. T. L. et al. Análise exploratória para a modelagem da frequência de acidentes de trânsito agregados ao nível de zonas de tráfego. **Transportes**, v. 23, n. 4, p. 42, 2015.

GOMES, M. J. T. L.; CUNTO, F.; SILVA, A. R. Geographically weighted negative binomial regression applied to zonal level safety performance models. **Accident Analysis & Prevention**, v. 106, n. June, p. 254–261, 2017.

GÓMEZ, F.; BOCAREJO, J. P. Accident Prediction Models for Bus Rapid Transit Systems. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2512, p. 38–45, 2015.

GREIBE, P. Accident prediction models for urban roads. **Accident Analysis and Prevention**, v. 35, n. 2, p. 273–285, 2003.

GUO, F.; WANG, X.; ABDEL-ATY, M. A. Modeling signalized intersection safety with corridor-level spatial correlations. **Accident Analysis and Prevention**, v. 42, n. 1, p. 84–92, 2010.

HADAYEGHI, A. **Use of Advanced Techniques to estimate Zonal Level Safety Planning Models and Examine Their Temporal Transferability**. Thesis. Department of Civil Engineering. University of Toronto, 2009.

HADAYEGHI, A.; SHALABY, A.; PERSAUD, B. Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. **Transportation Research Record: Journal of the Transportation Research Board**, v. 1840, p. 87–95, 24 jan. 2003.

HADAYEGHI, A.; SHALABY, A. S.; PERSAUD, B. N. Safety Prediction Models: Proactive Tool for Safety Evaluation in Urban Transportation Planning Applications. **Transportation Research Record**, v. 2019, n. 1, p. 225–236, 2007.

HAMZEHEI, A.; CHUNG, E.; MISKA, M. **Pre-crash traffic flow trend analysis on motorways**. International Symposium on Recent Advances in Transport Modelling. 2013.

HAQUE, M.; CHIN, H. Right-Angle Crash Vulnerability of Motorcycles at Signalized Intersections. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2194, p. 82–90, 2010.

HAUER, E. The engineering of safety and the safety of engineering. In: **Rothe, J. (Ed.), Challenging the Old Order: Towards New Directions in Traffic Safety Theory**. London: Transaction Publishers, 1990. p. 29–71. 1990.

HAUER, E. Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation. **Accident Analysis and Prevention**, v. 33, p. 799–808, 2001.

HAUER, E. Statistical Road Safety Modeling. **Transportation Research Record: Journal of the Transportation Research Board**, v. 1897, p. 81–87, 2004.

HAUER, E. Speed and Safety. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2103, p. 10–17, dez. 2009.

HAUER, E. **The Art of Regression Modeling in Road Safety**. Springer, 2015.

HAUER, E.; BAMFO, J. Two tools for finding what function links the dependent variable to the explanatory variables. **Proceedings of the ICTCT 1997 Conference**, n. October, p. 1–18, 1997.

HAUER, E.; HAKKERT, A. S. Extent and some implications of incomplete accident reporting. **Transportation Research Record**, v. 1185, p. 1–10, 1988.

HERMANS, E. et al. Benchmarking road safety: Lessons to learn from a data envelopment analysis. **Accident Analysis and Prevention**, v. 41, n. 1, p. 174–182, 2009.

HOJATI, A. T. et al. Modelling total duration of traffic incidents including incident detection and recovery time. **Accident Analysis and Prevention**, v. 71, p. 296–305, 2014.

HOSSAIN, M.; MUROMACHI, Y. Understanding crash mechanism on urban expressways using high-resolution traffic data. **Accident Analysis and Prevention**, v. 57, p. 17–29, 2013.

HOURDOS, J. et al. Real-Time Detection of Crash-Prone Conditions at Freeway High-Crash Locations. **Transportation Research Record**, v. 1968, n. 1, p. 83–91, 2006.

HUANG, H. et al. Macro and micro models for zonal crash prediction with application in hot zones identification. **Journal of Transport Geography**, v. 54, p. 248–256, 2016.

HUANG, Z. et al. Utilizing latent class logit model to predict crash risk. **Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017**, p. 161–165, 2017.

IMPRIALOU, M. I. M. et al. Re-visiting crash-speed relationships: A new perspective in crash modelling. **Accident Analysis and Prevention**, 2016.

JHAVERI, C. S.; PERRIN-JR, J.; MARTIR, P. T. SCOOT Adaptive Signal Control: An Evaluation of its Effectiveness over a Range of Congestion Intensities. In: TRANSPORTATION RESEARCH BOARD 82ND ANNUAL MEETING. **Anais...** Washington: TRB, 2003.

JIANG, X.; ADELI, H. Wavelet Packet-Autocorrelation Function Method for Traffic Flow Pattern Analysis. **Computer-Aided Civil and Infrastructure Engineering**, v. 19, n. 5, p. 324–337, 2004.

JONATHAN, A. V.; WU, K. F.; DONNELL, E. T. A multivariate spatial crash frequency model for identifying sites with promise based on crash types. **Accident Analysis and Prevention**, v. 87, p. 8–16, 2016.

JONSSON, T. et al. Differences in the Performance of Safety Performance Functions Estimated for Total Crash Count and for Crash Count by Crash Type. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2102, n. 1, p. 115–123, 2009.

JOSHUA, S. C.; GARBER, N. J. Estimating truck accident rate and involvements using linear and poisson regression models. **Transportation Planning and Technology**, v. 15, n. 1, p. 41–58, 1990.

JOVANIS, P. P.; CHANG, H.-L. Modeling the relationship of accidents to miles traveled. **Transportation Research Record 1068**, p. 45–51, 1986.

KIM, D. G.; WASHINGTON, S. The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. **Accident Analysis and Prevention**, v. 38, n. 6, p. 1094–1100, 2006.

KIM, S. S.; REDDY, A. L. N.; VANNUCCI, M. Detecting traffic anomalies using discrete wavelet transform. **Information Networking**, p. 951–961, 2004.

KUMARA, S. S. P.; CHIN, H. C. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. **Traffic Injury Prevention**, v. 4, n. 1, p. 53–57, 2003.



KWAK, H. C.; KHO, S. Predicting crash risk and identifying crash precursors on Korean expressways using loop detector data. **Accident Analysis and Prevention**, v. 88, p. 9–19, 2016.

LAMBERT, D. Zero-inflated poisson regression, with an application to defects in manufacturing. **Technometrics**, v. 34, n. 1, p. 1–14, 1992.

LEE, C.; ABDEL-ATY, M.; HSIA, L. Potential Real-Time Indicators of Sideswipe Crashes on Freeways. **Transportation Research Record**, v. 1953, n. 06, p. 41–49, 2006.

LEE, C.; HELLINGA, B.; SACCOMANNO, F. Real-Time Crash Prediction Model for Application to Crash Prevention in Freeway Traffic. **Transportation Research Record: Journal of the Transportation Research Board**, v. 1840, p. 67–77, 24 jan. 2003.

LEE, C.; SACCOMANNO, F.; HELLINGA, B. Analysis of Crash Precursors on Instrumented Freeways. **Transportation Research Record**, v. 1784, n. 1, p. 1–8, 2002.

LEE, J.; MANNERING, F. Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis. **Accident Analysis and Prevention**, v. 34, n. 2, p. 149–161, 2002.

LÍDER-DPVAT. **Relatório Anual. Seguradora Líder-DPVAT**. [s. l.]: [s. n.], 2017.

LINDEN, R. Técnicas de Agrupamento. **Revista de Sistemas de Informação da FSMA**, v. 4, p. 18–36, 2009.

LORD, D.; BONNESON, J. Development of Accident Modification Factors for Rural Frontage Road Segments in Texas. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2023, p. 20–27, 2007.

LORD, D.; MANAR, A.; VIZIOLI, A. Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. **Accident Analysis and Prevention**, v. 37, n. 1, p. 185–199, 2005.

LORD, D.; MANNERING, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. **Transportation Research Part A: Policy and Practice**, v. 44, n. 5, p. 291–305, 2010.

LORD, D.; MIRANDA-MORENO, L. F. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: A Bayesian perspective. **Safety Science**, v. 46, n. 5, p. 751–770, 2008.

LORD, D.; PERSAUD, B. Accident Prediction Models With and Without Trend: Application of the Generalized Estimating Equations Procedure. **Transportation Research Record: Journal of the Transportation Research Board**, v. 1717, n. 00, p. 102–108, 2000.

LORD, D.; WASHINGTON, S. P.; IVAN, J. N. Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. **Accident Analysis and Prevention**, v. 37, n. 1, p. 35–46, 2005.

LYON, C.; PERSAUD, B. Pedestrian collision prediction models for urban intersections. **Transportation Research Record: Journal of the Transportation Research Board**, v. d, n. 02, p. 102–107, 2002.

MA, J. Bayesian multivariate poisson-lognormal regression for crash prediction on rural two-lane highways. **The University of Texas at Austin**, 2006.

MA, J.; KOCKELMAN, K. M.; DAMIEN, P. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. **Accident Analysis and Prevention**, v. 40, n. 3, p. 964–975, 2008.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability**, v. 1, n. 233, p. 281–297, 1967.

MARTIN, J. L. Relationship between crash rate and hourly traffic flow on interurban motorways. **Accident Analysis and Prevention**, v. 34, n. 5, p. 619–629, 2002.

MOHAMMADI, M. A.; SAMARANAYAKE, V. A.; BHAM, G. H. Crash frequency modeling using negative binomial models: An application of generalized estimating equation to longitudinal data. **Analytic Methods in Accident Research**, v. 2, p. 52–69, 2014.

MONTAZERI-GH, M.; FOTOUHI, A. Traffic condition recognition using the k-means clustering method. **Scientia Iranica**, v. 18, n. 4 B, p. 930–937, 2011.

MOUNTAIN, L.; MAHER, M.; FAWAZ, B. The influence of trend on estimates of accidents at junctions. **Accident Analysis and Prevention**, v. 30, n. 5, p. 641–649, 1998.

NILSSON, G. **Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety**. [s.l.] Lund Institute of Technology, 2004.

OH, J.-S. et al. Real-Time Estimation of Accident Likelihood for Safety Enhancement. **Journal of Transportation Engineering**, v. 131, n. 5, p. 358–363, 2005.

PANDE, A.; ABDEL-ATY, M. Comprehensive Analysis of the Relationship Between Real-Time Traffic Surveillance Data and Rear-End Crashes on Freeways. **Transportation Research Record**, v. 1953, n. 1953, p. 31–40, 2006.

PARK, E. S.; LORD, D. Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. **Transportation Research Record**, v. 2019, n. 1, p. 1–6, 2008.

PEI, X.; WONG, S. C.; SZE, N. N. The roles of exposure and speed in road safety analysis. **Accident Analysis and Prevention**, v. 48, p. 464–471, 2012.

PERSAUD, B. N. Accident Prediction Models for Rural Junctions. **Canadian Journal of Civil Engineering**, v. 21, n. 4, p. 547–554, 1993.

PIRDAVANI, A. et al. Application of a Rule-Based Approach in Real-Time Crash Risk Prediction Model Development Using Loop Detector Data. **Traffic Injury Prevention**, v. 16,

n. 8, p. 786–791, 2015.

QIN, X.; IVAN, J. N.; RAVISHANKER, N. Selecting exposure measures in crash rate prediction for two-lane highway segments. **Accident Analysis and Prevention**, v. 36, n. 2, p. 183–191, 2004.

QUDDUS, M. A.; WANG, C.; ISON G, S. Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models. **Journal of Transportation Engineering**, v. 136, n. 5, p. pp 424-435, 2010.

RIBEIRO, E. R.; CUNHA, A. L. Análise exploratória de método utilizando Wavelet para detecção de padrões e anomalias em dados históricos do tráfego veicular. In: CONGRESSO NACIONAL DE PESQUISA EM TRANSPORTE ANPET - 2017, Recife. **Anais... ANPET**, 2017.

ROSHANDEL, S.; ZHENG, Z.; WASHINGTON, S. Impact of real-time traffic characteristics on freeway crash occurrence: Systematic review and meta-analysis. **Accident Analysis and Prevention**, v. 79, p. 198–211, 2015.

SAMANT, A.; ADELI, H. Feature Extraction for Traffic Incident Detection Using Wavelet Transform and Linear Discriminant Analysis. **Computer-Aided Civil and Infrastructure Engineering**, v. 15, n. 4, p. 241–250, 2000.

SCHULZE, H.; KOSSMANN, I. The role of safety research in road safety management. **Safety Science**, v. 48, n. 9, p. 1160–1166, 2010.

SHANKAR, V. N. et al. Modeling crashes involving pedestrians and motorized traffic. **Safety Science**, v. 41, n. 7, p. 627–640, 2003.

SHI, Q.; ABDEL-ATY, M. Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. **Transportation Research Part C: Emerging Technologies**, v. 58, p. 380–394, 2015.

SIDDIQUI, C.; ABDEL-ATY, M.; CHOI, K. Macroscopic spatial analysis of pedestrian and bicycle crashes. **Accident Analysis and Prevention**, v. 45, p. 382–391, 2012.

SILGU, M. A.; ÇELIKOĞLU, H. B. K-Means Clustering Method to Classify Freeway Traffic Flow Patterns. **Pamukkale University Journal of Engineering Sciences**, v. 20, n. 6, p. 232–239, 2014.

SOBREIRA, L. T. P.; CUNTO, F. J. C. Análise de Variações Temporais na Estimacão de Acidentes em Segmentos Arteriais Urbanos. **Transportes**, v. 22, p. 128–137, 2014.

SOBREIRA, L. T. P.; CUNTO, F. J. C. Análise Exploratória Espacial De Atropelamentos Em Zonas De Tráfego De Fortaleza. In: CONGRESSO NACIONAL DE PESQUISA EM TRANSPORTE ANPET - 2017, Recife. **Anais... ANPET**, 2017.

SOLOMON, D. . Crashes on main rural highways related to speed, driver and vehicle. **Bureau of Public Roads**, 1964.

STEMPFEL, J. et al. Effects of urban congestion on safety of networks. **Journal of Transportation Safety & Security**, v. 8, n. 3, p. 214–229, 2 jul. 2016.

STUTER, J. **Aggressive Driving Enforcement, DOT HS 809 707**. Washington, DC.

TANISHITA, M.; VAN WEE, B. Impact of vehicle speeds and changes in mean speeds on per vehicle-kilometer traffic accident rates in Japan. **IATSS Research**, v. 41, n. 3, p. 107–112, 2017.

THANG, T. M.; KIM, J. The anomaly detection by using DBSCAN clustering with multiple parameters. **2011 International Conference on Information Science and Applications, ICISA 2011**, 2011.

THEOFILATOS, A. et al. Impact of real-time traffic characteristics on crash occurrence: Preliminary results of the case of rare events. **Accident Analysis and Prevention**, n. May 2016, p. 1–9, 2018.

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. **Journal of the Royal Statistical Society. Series B: Statistical Methodology**, v. 63, n. 2, p. 411–423, 2001.

TORRES, C. A. **Avaliação do Desempenho da Segurança Viária em Redes Urbanas de Transportes com Utilização de Modelos de Previsão de Acidentes**. Dissertação (Mestrado em Engenharia de Transportes) - Departamento de Engenharia de Transportes, Universidade Federal do Ceará, Fortaleza, 2016.

UKKUSURI, S. et al. The role of built environment on pedestrian crash frequency. **Safety Science**, v. 50, n. 4, p. 1141–1151, 2012.

USDOT. **Intelligent Transportation Systems (ITS) Strategic Plan 2015-2019**. [s.l: s.n.], 2014.

WANG, C.; QUDDUS, M. A.; ISON, S. G. Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England. **Accident Analysis and Prevention**, v. 41, n. 4, p. 798–808, 2009.

WANG, X. et al. Speed, speed variation and crash relationships for urban arterials. **Accident Analysis and Prevention**, v. 113, n. November 2016, p. 236–243, 2018.

WASHINGTON, S. et al. **Incorporating safety into long-range transportation planning** National Cooperative Highway Research Program. Washington, DC. 2006.

WEIJERMARS, W.; VAN BERKUM, E. Analyzing highway flow patterns using cluster analysis. **Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.**, n. October, p. 831–836, 2005.

WHO - World Health Organization. **Global status report on road safety 2015**. Geneva: [s. n.], 2015.

WHO - World Health Organization. **Managing Speed**. Geneva: [s. n.], 2017.

WONG, S. C.; SZE, N. N.; LI, Y. C. Contributory factors to traffic crashes at signalized intersections in Hong Kong. **Accident Analysis and Prevention**, v. 39, n. 6, p. 1107–1113, 2007.

WU, W. Q. et al. Application of generalized estimating equations for crash frequency modeling with temporal correlation. **Journal of Zhejiang University-Science A**, v. 15, n. 7, p. 529–539, 2014.

XIE, K. et al. Corridor-level signalized intersection safety analysis in Shanghai, China using Bayesian hierarchical models. **Accident Analysis and Prevention**, 2013.

YANG, B. Z.; LOO, B. P. Y. Land use and traffic collisions: A link-attribute analysis using Empirical Bayes method. **Accident Analysis and Prevention**, v. 95, p. 236–249, 2016.

YANG, X.; LU, Y. C.; LIN, Y. Optimal Variable Speed Limit Control System for Freeway Work Zone Operations. **Journal of Computing in Civil Engineering**, v. 31, n. 1, 2016.

ZHENG, Z. et al. Applications of wavelet transform for analysis of freeway traffic: Bottlenecks, transient traffic, and traffic oscillations. **Transportation Research Part B: Methodological**, v. 45, n. 2, p. 372–384, 2011.

ZHENG, Z. Empirical Analysis on Relationship between Traffic Conditions and Crash Occurrences. **Procedia - Social and Behavioral Sciences**, v. 43, p. 302–312, 2012.

ZHENG, Z.; AHN, S.; MONSERE, C. M. Impact of traffic oscillations on freeway crash occurrences. **Accident Analysis and Prevention**, v. 42, n. 2, p. 626–636, 2010.

ZHOU, M.; SISIOPIKU, V. Relationship Between Volume-to-Capacity Ratios and Accident Rates. **Transportation Research Record**, v. 1581, n. 1, p. 47–52, 1997.