

UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA
DEPARTAMENTO DE TRANSPORTES

OLÍMPIO MENDES DE BARROS

**ESTIMAÇÃO DAS CARACTERÍSTICAS DE TRÁFEGO EM TEMPO PRÓXIMO AO REAL PARA USO
EM SISTEMAS DE PREVISÃO E CONTROLE DE TRÁFEGO**

São Paulo
2022

OLÍMPIO MENDES DE BARROS

**ESTIMAÇÃO DAS CARACTERÍSTICAS DE TRÁFEGO EM TEMPO PRÓXIMO AO REAL PARA USO
EM SISTEMAS DE PREVISÃO E CONTROLE DE TRÁFEGO**

VERSÃO ORIGINAL

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção do
título Doutor em Ciências.

Área de Concentração:
Engenharia de Transportes

Orientador: Prof. Dr. Edvaldo Simões da
Fonseca Júnior

São Paulo
2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na publicação
Serviço de Biblioteca e Documentação
Escola Politécnica da Universidade de São Paulo

Barros, Olímpio Mendes de
ESTIMAÇÃO DAS CARACTERÍSTICAS DE TRÁFEGO EM TEMPO
PRÓXIMO AO REAL PARA USO EM SISTEMAS DE PREVISÃO E CONTROLE
DE TRÁFEGO / O. M. Barros -- São Paulo, 2022.
160 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo.
Departamento de Engenharia de Transportes.

1. Engenharia de Transportes 2. Dados em Painel 3. Tráfego Urbano
4. Previsão de Tráfego I. Universidade de São Paulo. Escola Politécnica.
Departamento de Engenharia de Transportes II. t.

Nome: Barros, Olímpio Mendes de

Título: Estimação das características de tráfego em tempo próximo ao real para o uso em sistemas de previsão e controle de tráfego

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção do
título Doutor em Ciências.

Aprovado em:

Banca Examinadora

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Prof. Dr. _____
Instituição: _____
Julgamento: _____

Dedicatória

À minha amada esposa e companheira Silvia Carla Itaborai, pelo amor e carinho e pelos conselhos e apoio incondicionais durante toda esta caminhada.

Ao meu filho Paulo Vitor, que aprendeu a ler durante meu mestrado e agora, mestrando, tem sido um companheiro de pesquisa e ciência.

Agradecimentos

Agradeço à Deus, pela saúde e capacidade em desenvolver este trabalho até o fim.

À minha esposa Silvia, pelo amor, carinho, companhia, pelos inúmeros mimos durante o desenvolvimento desta pesquisa, pela compreensão das muitas horas de dedicação aos estudos e apoio incondicional na caminhada.

Ao meu filho Paulo Vitor, pelo carinho, conselhos, companhia e valoroso apoio na parte científica e computacional. Por entender minhas ausências durante os longos trabalhos de pesquisa.

Aos meus pais, in memoriam, pelo amor ao longo da vida e pelo direcionamento aos estudos.

À minha irmã e meus irmãos, cunhado, cunhadas, sobrinhos e sobrinhas e demais familiares que sempre me apoiaram e incentivaram.

Aos meus orientadores Professores Doutores Edvaldo Simões da Fonseca Júnior e Claudio Luiz Marte, pela orientação e apoio, com importantes direcionamentos e correções ao longo desta jornada.

Aos alunos de graduação da Poli/USP: Álvaro Miguel Linares, Caio Gonçalves de Moraes, Caio Ubaldo Neri, Domenico Zemma, Emerson Fernandes Sabino, Guilherme Mazzuia Stocco, Kaiene Domingues Paz, Marcella Gomes da Costa Pereira, Marco Aurélio Tavares Marques da Silva, Mariane Ayumi Bando, Michelle Mieko Taniguchi Trintinalia, Murilo Yoshio Matsubara, Ricardo Bonfiglioli Junqueira Andrade, Rodrigo Pina Bluma e Yuri Gil Ferreira, que, através de seus trabalhos de final de curso e estágios de iniciação científica, colaboraram de modo significativo no desenvolvimento desta tese.

Aos professores e colegas da Poli, FEA e Arquitetura da USP, com quem tive a oportunidade de conviver e aprender ao longo destes últimos anos.

Aos colegas da CET pelo incentivo, apoio e companheirismo nesta caminhada.

À SMMT – Secretaria Municipal de Mobilidade e Trânsito de São Paulo e à CET, pela liberação de dados e pela liberação para frequentar o curso e desenvolver as pesquisas necessárias

Epígrafe

“Céptico como os cépticos, crente como os crentes.

A metade que avança é crente, a metade que confirma é céptica.

Mas o cientista perfeito é também jardineiro: acredita que a beleza é conhecimento.”

Gonçalo Tavares – Breves Notas Sobre Ciência

RESUMO

BARROS, Olímpio Mendes de. **Estimação das características de tráfego em tempo próximo ao real para o uso em sistemas de previsão e controle de tráfego**. 2021. 136 f. Texto para Defesa de Doutorado – Departamento de Transportes, Escola Politécnica, Universidade de São Paulo, São Paulo, 2019.

O tráfego de veículos no ambiente urbano se apresenta complexo, com variações que dependem de seus usuários, de condições meteorológicas e de situações que impactam o fluxo através da rede viária. Equipamentos de fiscalização, distribuídos pelas vias, podem ser utilizados como sensores para outros sistemas de tráfego. Existe a necessidade de estudos a respeito de métodos que possam tratar os dados, gerados por esses equipamentos, transformando-os em informações úteis para os sistemas de monitoramento e controle das condições do tráfego. A estimação de tráfego consiste em atualizar, para o momento presente, as informações de tráfego defasadas - recebidas de sensores e outras fontes - tendo em vista os tempos de transmissão, tratamento e processamento dos dados. Assim, a estimação pode ser feita olhando-se para um intervalo à frente ao da coleta dos dados. O objetivo desta tese é propor uma metodologia para estimar as características espaço-temporais do tráfego em condições dinâmicas (tempo próximo ao real), visando a previsão de tráfego de curto prazo, para uso em sistemas de controle e informação de tráfego. A metodologia de estimação proposta nesta pesquisa inclui, entre outros, os passos de montagem de séries temporais e aplicação de modelos estatísticos para a estimação das condições de tráfego. Para o processo de estimação foram aplicados modelos de “Autorregressão Espaço-Temporal de Médias Móveis Integradas” (STARIMA) e “Regressões para Dados em Painel”. No caso dos dados em painel, verificou-se a aplicação de Modelos Longitudinais e Modelos Lineares Hierarquizados (HLM) pelo local - via e sensor - e pelo tempo - dia útil, dia da semana e período do dia. Diversas variáveis explicativas foram testadas nos modelos em painel, incluindo a quantidade de chuva, presença de faixa exclusiva de ônibus, quantidade de faixas de rolamento e classe da via. Para aplicação do modelo STARIMA foi elaborada uma pesquisa detalhada sobre as matrizes de peso espacial. Como resultado, foi verificada a viabilidade de uso de matrizes agrupadas, que são mais simples em comparação com os outros modelos utilizados e inédito nos modelos de estimação e previsão de tráfego com STARIMA. Matrizes com pesos assimétricos, baseados no tempo ou velocidade de percurso superaram os modelos simétricos que são mais usuais nas previsões de volume (fluxo). As previsões de velocidade pontual, feitas com cenários baseados em Dados em Painel, obtiveram acurácia um pouco superior ao STARIMA nos períodos com tráfego normal. Para os períodos de tráfego atípico o STARIMA apresentou melhor acurácia. Os modelos testados se mostraram viáveis aos processos de estimação de tráfego em tempo próximo ao real, com MAPE entre 6% e 7%, tanto nas estimações de volume como de velocidade. Esta tese apresenta a aplicação de dados reais e detalha as etapas e critérios de desenvolvimento dos modelos, de modo a possibilitar que outros pesquisadores apliquem os modelos em situações reais e em futuras pesquisas.

Palavras chave: Estimação de tráfego, Previsão de tráfego, Tráfego urbano, STARIMA, HLM, Dados em painel, Matriz W, Matriz de correlação

Abstract

Vehicle traffic in the urban environment is complex, containing variations that depend on its users, weather conditions, and situations that impact the flow through the road network. Traffic surveillance equipment distributed along the roads can be used as sensors for other traffic systems. There is a need for studies on methods that can treat the data generated by this equipment, transforming it into useful information for traffic monitoring systems. Traffic estimation consists in updating, for the present moment, the lagged traffic information - received from sensors and other sources - considering the data transmission, treatment and processing times. Thus, estimation can be done by looking at a period ahead of that of the data collection. The objective of this thesis is to propose a methodology to estimate the space-time characteristics of traffic under dynamic conditions (near real-time), for daily operation of urban traffic, aiming at short-term traffic forecasting. The estimation methodology proposed in this research includes, among others, time series assembly steps and application of statistical models for the estimation of traffic conditions. For the estimation process, "Space-Time Autoregression Integrated Moving Averages" (STARIMA) and "Regressions for Panel Data" models were applied. In the case of panel data, Longitudinal Models and Hierarchical Linear Models (HLM) were applied by location - track and sensor - and by time - weekday, weekday, and period of the day. Several explanatory variables were tested in the panel models, including the amount of rain, the presence of exclusive bus lanes, the number of lanes, and the class of roadway. For the application of STARIMA, a detailed research on spatial weight matrices was carried out. As a result, the feasibility of using clustered matrices was verified, considering that they are simpler compared to the other models used and unprecedented in traffic estimation and prediction models with STARIMA. Matrices with asymmetric weights based on travel time or speed outperformed the symmetric models that are more usual in volume (flow) predictions. The instantaneous speed predictions made with scenarios based on Panel Data obtained slightly better accuracy than the best scenario with STARIMA. The point speed predictions, made with scenarios based on Panel Data, obtained slightly better accuracy than STARIMA in periods with regular traffic. For periods of atypical traffic, STARIMA showed better accuracy. The tested models proved to be feasible for traffic estimation processes in near real time, with MAPE between 6% and 7%, both in volume and speed estimations. This thesis presents the application of real data and details the steps and criteria of model development to enable other researchers to apply the models in real situations and in future research.

Keywords: traffic estimation, traffic forecasting, urban traffic, STARIMA, HLM, panel data, W matrix, correlation matrix

LISTA DE FIGURAS

Figura 1-1 – Fluxograma da metodologia proposta.....	17
Figura 2-1 – Métodos de Previsão de Tráfego de Curto Prazo (STTF).....	22
Figura 2-2 – Ordem de vizinhanças unidirecionais utilizadas em matrizes espaciais	30
Figura 2-3 - Flutuação horária das viagens por motivo (a) e por modo (b)	39
Figura 2-4 – Medição de congestionamento horário (a) e diário (b) em todo Município de São Paulo	40
Figura 3-1 – Fluxograma da metodologia proposta.....	42
Figura 3-2 – Gráficos STPACF e ATACF	51
Figura 3-3 – Exemplo de estrutura hierárquica por local	53
Figura 3-4 – Rede viária teórica.....	58
Figura 3-5 - Matriz de Peso Geral da Rede Teórica com pesos do tipo 1-1/V.....	59
Figura 3-6 - Matriz de Ordens de Vizinhança por Contiguidade	59
Figura 3-7 - Matrizes geradoras das matrizes de peso por ordem para vizinhança por contiguidade..	60
Figura 3-8 – Exemplo de Matriz de peso da Rede Teórica com vizinhança por contiguidade	60
Figura 3-9 - Matriz de Ordens de Vizinhança baseada no modelo LP	61
Figura 3-10 - Exemplo de Matriz de peso da Rede Teórica com vizinhança LP	62
Figura 3-11 - Matriz de Pesos Agrupada da Rede Teórica com pesos do tipo 1-1/V	62
Figura 4-1 - Mapa com a indicação dos radares e links do estudo	68
Figura 4-2 – Características do Tráfego	73
Figura 4-3 - Distribuição do Fluxo em 10 radares nos dois dias de treino e teste do modelo STARIMA	74
Figura 4-4 – Fluxo horário por tipo de veículo para um dia típico	75
Figura 4-5 – Porcentagem de veículos identificados ao longo do dia, de acordo com o tipo e veículo	76
Figura 4-6 – Exemplo da série temporal pontual dos volumes para uso com STARIMA.....	80
Figura 4-7 – Detalhe do banco de dados utilizado	87
Figura 5-1 – Matriz Geral de Pesos 1-1/T utilizada nos testes.....	95
Figura 5-2 – Característica dos dados de treino originais e estacionários após diferenciação	96
Figura 5-3 – Processo de Rolagem Iterativa de Dados para Estimação com STARIMA	98
Figura 5-4 – Acurácia de Previsão dos pesos testados nos corredores em separado e com toda a rede - STARIMA	101
Figura 5-5 – Dados aninhados por local - Eixo e Sensor	105
Figura 5-6 – Dados aninhados por tempo – Dia da Semana e Período do dia	106
Figura 5-7 – Precisão da Previsão de acordo com a Vizinhança e Característica do Tráfego	108
Figura 5-8 – Acurácia de Previsão de acordo com o tipo de Vizinhança e Horizonte de Previsão	109
Figura 5-9- Acurácia de previsão dos pesos testados em separado e na rede	111
Figura 5-10 – Gráfico da previsão com vários horizontes.....	112

Figura 5-11 –Acurácia de Acordo com a Característica do Tráfego e o Horizonte de Previsão	113
Figura 5-12 – Estimação, Previsão de 20min e Velocidade Pontual Média Medida	114
Figura 5-13 – Estimação das Características de Tráfego – STARIMA 1/D	116
Figura 5-14 – Modelo Nulo Aninhado por Local	119
Figura 5-15 – Modelo Nulo Aninhado por Tempo	120
Figura 5-16 – Modelos Aninhados por Local com Efeitos Aleatórios.....	122
Figura 5-17 – Modelo Longitudinal de Dados em Painel	125
Figura 5-18 – Modelos HLM e Efeitos Aleatórios de Tempo Sem e Com Autocorrelação	128
Figura 5-19 – Comparação dos Tipos de Estimação no ambiente de teste	131

LISTA DE TABELAS

Tabela 2-1- Pesos e Vizinhanças em ST-STTF com STARIMA	31
Tabela 3-1 – Organização dos dados para aplicação em modelos baseados em STARIMA	50
Tabela 3-2 – Organização dos dados para aplicação em modelos Longitudinais com Dados em Painel	52
Tabela 3-3 – Organização dos dados para aplicação em modelos Hierárquicos com Dados em Painel	54
Tabela 3-4 - Características das matrizes utilizadas nos testes	56
Tabela 3-5 – Tempo de percurso e Ordem de Vizinhança LP nos trechos do exemplo teórico	61
Tabela 4-1 – Características dos Trechos do estudo.....	69
Tabela 4-2 – Dados originais dos sensores do tipo radar da SMMT-SP	71
Tabela 4-3 - Fatores de equivalência por UCP	79
Tabela 4-4 – Características dos dados pontuais.....	81
Tabela 4-5 – Características dos dados de trecho	85
Tabela 4-6 – Variáveis que consideram dados pontuais e de trecho	86
Tabela 5-1 – Exemplo de calibração exaustiva de um modelo	97
Tabela 5-2–Recorte dos resultados de testes feitos com toda a Rede – RMSE	99
Tabela 5-3 – Análise de Acurácia dos Modelos Testados – MAPE	102
Tabela 5-4 – Análise de Acurácia dos Modelos Testados - RMSE	103
Tabela 5-5 – Porcentagem de vezes em que o modelo obteve melhor acurácia de acordo com a Vizinhança	108
Tabela 5-6 – Comparação da Acurácia por Tipo de Peso (MAPE)	110
Tabela 5-7 – Desempenho dos modelos STARIMA na estimação da velocidade pontual	115
Tabela 5-8 – Desempenho dos modelos Nulos	117
Tabela 5-9 – Desempenho dos modelos com efeitos aleatórios	121
Tabela 5-10 – Desempenho dos modelos longitudinais	124
Tabela 5-11 – Desempenho dos modelos longitudinais	127
Tabela 5-12 – Comparação da acurácia dos modelos de estimação de velocidade para os ambientes de teste	129

LISTA DE ABREVIÇÕES

AG – Vizinhaça Agrupada	SD-STARIMA – Seasonal Difference Space-Time
AIC – <i>Akaike Information Criterion</i> - Critério de Informação de Akaike	<i>Autoregressive Integrated Moving Average</i> – STARIMA com diferenciação sazonal
AR - Autorregressão	SMMT-SP - Secretaria Municipal de Mobilidade e Transportes do Município de São Paulo
ARIMA - <i>Autoregressive Integrated Moving Average</i> – Autorregressão de Médias Móveis Integradas	SPTRans – São Paulo Transportes
CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior	ST – <i>Space-Time</i> - Espaço-temporal
CET-SP - Companhia de Engenharia de Tráfego de São Paulo	STACF - <i>Space Time Autocorrelation Function</i> – Função de Autocorrelação Espaço-Temporal
CoD - Coeficiente de Determinação	STAR – <i>Space Time Autoregressive</i> - Autorregressão Espaço-Temporal
csv – extensão de arquivo do tipo texto separado por vírgulas	STARIMA – <i>Space Time Autoregressive Integrated Moving Average</i> - Autorregressão Espaço-Temporal de Médias Móveis Integradas
DTA - <i>Dynamic Traffic Assignment</i> - Alocação Dinâmica do Tráfego	STARIMAR - <i>Space-Time Autoregressive Integrated Moving Average Regressive</i> – Autorregressão Espaço-Temporal de Médias Móveis Regressivas
GARCH – <i>Generalized autoregressive conditional heteroscedasticity</i> - heteroscedasticidade condicional autorregressiva generalizada	STPACF - <i>Space Time Partial Autocorrelation Function</i> – Função de Autocorrelação Parcial Espaço-Temporal
GBRT - <i>Gradient Boosting Regression Tree</i> – Árvore de Regressão com Potenciação de Gradiente	ST-STTF - STTF espaço temporal
GESD - <i>Generalized Extreme Studentized Deviate</i> – Desvio Padrão Extremo Generalizado	STTF - <i>Short Time Traffic Forecast</i> – Previsão de Tráfego de Curto Prazo
GLMM – <i>Generalized Linear Multilevel Model</i> – Modelo Multinível Linear Generalizado	txt – extensão de arquivo do tipo texto
GPS – <i>Global Positioning System</i> – Sistema de Posicionamento Global	UCP – Unidades de carros de passeio
GSTARIMA - <i>Generalised STARIMA</i> – STARIMA Generalizado	UN – Peso Unitário
HLM – <i>Hierarchical Linear Model</i> – Modelo Hierárquico Linear	VAR – <i>Vector Autoregressive</i> – Vetor Autorregressivo
ID – valor de identificação do veículo após a anonimização	VARMA - <i>Vector Autoregressive Moving Average</i> – Vetor Autorregressivo de Média Móvel
ITS - Intelligent Transportat Systems	VC - Vizinhança por Contiguidade
KNN - <i>k Nearest Neighbor</i>	ZIB - Regressão Binomial Inflada de Zeros
LAI – Lei de Acesso à Informação	ZIP - Regressão Poisson Inflada de Zeros
LP – <i>Lag</i> de Percurso	1/D = modelo de matriz com pesos obtidos pelo inverso da distância
LSTARIMA - <i>Localised STARIMA</i> – STARIMA Local	1/L = modelo de matriz com pesos obtidos pelo inverso do número de LAGs de percurso
MA – Média Móvel	1/T = modelo de matriz com pesos obtidos pelo inverso do tempo de percurso
MAPE – Mean Absolute Percentage Error – Erro Percentual Absoluto Médio	1/V = modelo de matriz com pesos obtidos pelo inverso da velocidade pontual
NA – Registro vazio, sem dado	1-1/L = modelo de matriz com pesos obtidos pelo inverso do número de LAGs de percurso e seu complemento
OD - origem destino	1-1/T = modelo de matriz com pesos obtidos pelo inverso do tempo de percurso e seu complemento
R – Linguagem de Programação R	1-1/V = modelo de matriz com pesos obtidos pelo inverso da velocidade e seu complemento
RFID - identificação por radiofrequência	
RMSE – Root Mean Square Error- Raiz do Erro Quadrático Médio	
SARIMA – <i>Seasonal Autoregressive Integrated Moving Average</i> – ARIMA Sazonal	

LISTA DE SÍMBOLOS

a = índice	q = ordem temporal da média móvel
c = constante	R_i = Radares
D = distância, extensão do trecho	r = um radar específico
d = ordem de diferenciação da série temporal	S = Desvio Padrão
f = índice	s = índice
G = Coeficiente de Grubbs	T = Tempo de Percurso no trecho
g = índice	t = instante em que é feita a medição
H = horizonte de previsão	v_a = valor atípico
h = diferença entre o terceiro e o primeiro quartis	v_e = valor extremo
i = índice	V = Velocidade
j = intervalo avaliado	V_p = Velocidade Pontual
k = quantidade de intervalos (lag temporal)	x = número de intervalos k necessários para percorrer o trecho a 4km/h
L = operador espacial	X = vetor de variáveis explicativas
\hat{L} = máximo valor da função de verossimilhança	Y = variável dependente
l = ordem de vizinhança	y = cada amostra da variável Y
l_{LP} = Ordem de vizinhança para o modelo LP	W = matriz de pesos espaciais
m = ordem espacial da média móvel	w = elemento da matriz W
N = quantidade de locais de estudo	z = identificador da contagem analisada
n = tamanho da amostra	α = parâmetro do modelo
nj = local à jusante	β = parâmetro do modelo
nm = local à montante	γ = parâmetro do modelo
O = Valor Observado	δ = parâmetro do modelo
\bar{O} = Valor Observado Médio	ε = erro estatístico
P = valor Previsto	λ = ordem espacial da autorregressão do modelo STARIMA e fator dos modelos ZIP e ZIB
p = ordem temporal da autorregressão	μ = média da variável Y
Q = Volume de veículos	χ^2 = Qui quadrado estatístico
Qt = Quartil	

Sumário

1	INTRODUÇÃO	15
1.1	MOTIVAÇÃO	18
1.2	OBJETIVO.....	18
1.3	QUESTÕES DE PESQUISA	19
1.4	ORGANIZAÇÃO	19
2	REVISÃO DA LITERATURA	21
2.1	PREVISÃO DE TRÁFEGO	21
2.1.1	Estimação do estado do tráfego ou previsão de tráfego?	24
2.2	ESTIMAÇÃO E PREVISÃO DE TRÁFEGO ESPAÇO-TEMPORAL.....	25
2.2.1	Estimação com uso de modelos autorregressivos espaço-temporais.....	26
2.2.2	Estimação com uso de regressão para dados em painel	27
2.2.3	Estimação com uso de modelos de simulação.....	29
2.3	MATRIZ DE PESO ESPACIAL.....	29
2.3.1	Pesos da matriz na ST-STTF.....	31
2.3.2	Vizinhanças da matriz na ST-STTF	32
2.4	MODELOS DE CONTROLE PREDITIVO	34
2.5	DADOS DE TRÁFEGO.....	35
2.5.1	Fontes de dados dinâmicos de tráfego	36
2.5.2	Parâmetros de caracterização do tráfego.....	37
2.5.3	Intervalos de agrupamento dos dados	38
2.6	SAZONALIDADE	38
2.7	CONSIDERAÇÕES FINAIS DO CAPÍTULO 2	40
3	METODOLOGIA.....	42
3.1	MÓDULOS COMPONENTES DA METODOLOGIA	43
3.1.1	Módulo de recepção e pré-processamento dos dados.....	43
3.1.2	Módulo de montagem das séries temporais	45
3.1.3	Módulo de estimação	46
3.1.4	Módulo de envio de dados a outros sistemas	48
3.2	MODELOS DE ESTIMAÇÃO	48
3.2.1	Modelo autorregressivo - STARIMA.....	49
3.2.2	Modelos de regressão de Dados em Painel	52
3.3	MATRIZES DE PESO ESPACIAL	55
3.3.1	Tipos de pesos utilizados nos modelos	57
3.3.2	Organização das matrizes de peso de acordo com as vizinhanças	58
3.4	ANÁLISE DOS DADOS DO TRÁFEGO	63

3.4.1	Valores atípicos - outliers	63
3.4.2	Dados ausentes	64
3.5	MEDIDAS DE DESEMPENHO E ACURÁCIA	65
3.6	CONSIDERAÇÕES FINAIS DO CAPÍTULO 3	66
4	PREPARAÇÃO DOS DADOS DA PESQUISA	67
4.1	LOCAL DA APLICAÇÃO DA PESQUISA	67
4.2	DADOS DE SENSORES	69
4.2.1	Dados disponíveis	70
4.2.2	Dados utilizados.....	72
4.2.3	Visão preliminar das características do tráfego	73
4.2.4	Identificação dos veículos.....	75
4.2.5	Dados de chuva	77
4.3	PRÉ-PROCESSAMENTO DOS DADOS	77
4.4	MONTAGEM DAS SÉRIES TEMPORAIS	78
4.4.1	Dados pontuais.....	79
4.4.2	Dados de trecho	83
4.4.3	Consolidação dos dados	86
4.5	ANÁLISE DE DADOS	88
4.5.1	Teste GESD	89
4.5.2	Teste do quartil.....	90
4.6	CONSIDERAÇÕES FINAIS DO CAPÍTULO 4	90
5	APLICAÇÃO E RESULTADOS	92
5.1	APLICAÇÃO	92
5.1.1	Estimação das características do tráfego com STARIMA	93
5.1.1.1	Análise das matrizes de relação espacial – variável volume	94
5.1.2	Estimação das Características do Tráfego com Dados em Painel	104
5.2	DISCUSSÃO DOS RESULTADOS.....	107
5.2.1	Estimação dos modelos com modelo STARIMA.....	107
5.2.1.1	Vizinhança	107
5.2.1.2	Pesos.....	109
5.2.1.3	Abrangência.....	111
5.2.1.4	Horizonte de previsão.....	112
5.2.1.5	Característica do tráfego no período	113
5.2.1.6	Modelos STARIMA para a característica velocidade pontual.....	114
5.2.2	Estimação com modelo de Dados em Painel	117
5.2.2.1	Modelos hierárquicos nulos	117

5.2.2.2	Análise das variáveis aleatórias	121
5.2.2.3	Modelos longitudinais	123
5.2.2.4	Modelos hierárquicos completos	126
5.2.3	Comparando os modelos de estimação da velocidade	129
5.3	CONSIDERAÇÕES FINAIS DO CAPÍTULO 5	131
6	CONCLUSÕES.....	133
	BIBLIOGRAFIA.....	136
	APÊNDICES	145
A.	Trabalhos relacionados e decorrentes:	145
B.	Aplicação de Dados em Painel na predição de volumes.....	148
C.	Scripts.....	150
D.	Aplicativo STARIMA no R Shiny	151
E.	Detecção de incidentes	152
a.	Regressão multivariada inflada de zeros	154
b.	Dados de ocorrências	155
i.	Dados do órgão de trânsito	155
ii.	Dados de aplicativo de roteamento	156

1 INTRODUÇÃO

A evolução dos *Intelligent Transport Systems* (ITS), e seu uso em Cidades Inteligentes, vem em resposta às novas demandas decorrentes do constante crescimento dos núcleos urbanos, que tem ocorrido nas últimas décadas.

Gestores de cidades necessitam tomar decisões rápidas diante de situações complexas e para isso carecem de ferramentas de gestão robustas, inteligentes e confiáveis. Ferramentas de gestão que tenham a característica de prever, ainda que de modo aproximado, as consequências de uma decisão, são de extrema relevância à gestão da mobilidade e devem se tornar cada vez mais presentes nas Centrais de Operação e Gestão das Cidades.

Nessa direção, surge a necessidade de sistemas que possam estimar a situação do tráfego, fornecendo essa informação diretamente aos gestores e se integrando a outros sistemas de gestão e previsão de curto prazo. Como exemplo, as ferramentas de estimação e previsão das condições de tráfego podem ser utilizadas em sistemas de controle de tráfego como cancelas e semáforos (OTOSHI et al., 2018).

Na mesma direção, a integração com sistemas de roteamento, pode propiciar ganhos coletivos. A elaboração de rotas sem considerar a situação futura do tráfego pode resultar na utilização de trajetos que estarão congestionados alguns minutos depois. A análise da utilização de um algoritmo, baseado na previsão do tráfego para o roteamento centralizado e dinâmico de veículos, indicou a redução de 50% nos tempos médios de viagem, caso todos os veículos utilizassem o algoritmo, em contraposição ao aumento de aproximadamente 80% caso todos os veículos utilizassem rotas determinadas de forma individual e sem incluir a previsão de tráfego (KIM et al., 2016).

Os estudos relacionados à estimação e previsão de tráfego ganharam força na última década, o que se comprova com a quantidade de Revisões Científicas (*Review*) publicadas sobre diversos aspectos desse tema:

- a) previsão de tráfego - avanços e desafios (LAÑA et al., 2018; VLAHOGIANNI; KARLAFTIS; GOLIAS, 2014);
- b) previsão de tráfego espaço-temporal (ERMAGUN; LEVINSON, 2018; GANG; SHOUHUI; XIAOBO, 2016; PAVLYUK, 2019);
- c) previsão de tráfego em *smart cities* (TASCIKARAOGLU, 2018; XIONG et al., 2012);

- d) previsão de tráfego em rodovias (OH et al., 2017);
- e) locação de sensores para previsão de tráfego (CASTILLO et al., 2015; GENTILI; MIRCHANDANI, 2018);
- f) estimação da densidade de tráfego (T. DARWISH; K. ABU BAKAR, 2015);
- g) alocação dinâmica de tráfego (AGHAMOHAMMADI; LAVAL, 2018; MITSAKIS; SALANOVA; GIANOPOULOS, 2011; WANG et al., 2018);
- h) métodos estatísticos e redes neurais na previsão de tráfego (DE LUCA; GALLO, 2017; KARLAFTIS; VLAHOIANNI, 2011);
- i) previsão de tráfego com uso de modelos autorregressivos (PAVLYUK, 2017; SCHIMBINSCHI et al., 2017);
- j) inteligência artificial e *big data* em sistemas de tráfego (VLAHOIANNI, 2015);
- k) uso de simuladores de tráfego em tempo real (PELL; MEINGAST; SCHAUER, 2017)

A análise destas revisões indica que a maior parte do material científico publicado está relacionado às análises pontuais e a pequenos trechos de vias expressas. Boa parte sequer considera a questão espacial, ficando direcionada apenas na relação temporal. Apenas mais recentemente verificou-se o surgimento de estudos em redes urbanas. Ainda assim, a grande maioria aborda estudos de pequenas redes ou com características específicas. Estas revisões também apresentam uma concentração de estudos baseados em dados históricos, com a aplicação de ferramentas de Inteligência Artificial e de análise de dados, os quais serão classificados mais adiante como métodos de Reconhecimento de Padrões.

Como será apresentado de modo mais detalhado ao longo dos próximos capítulos, os modelos de previsão dinâmica (em tempo próximo ao real) necessitam receber dados atualizados das condições do fluxo. Neste caso, se incluem tanto os sistemas de simulação de tráfego como os demais modelos de previsão. As condições do fluxo são determinadas com base em características (parâmetros) como: volume, velocidade, ocupação e tempos de percurso. Essas características podem ser extraídas de dados de sensores, instalados nas vias, e de fontes massivas de dados, aquelas provenientes de redes sociais e do monitoramento do posicionamento de veículos e telefones celulares.

Os sistemas de controle de tráfego, em especial os sistemas de controle semafórico, que trabalham em tempo real, necessitam igualmente de dados atualizados e organizados para efetuar os ajustes internos. Esses sistemas são baseados em Modelos de Controle Preditivo, e

têm como módulo inicial um processo de estimação das características do tráfego. A estimação pode ser feita com base em diversos métodos, como os utilizados nesta tese.

A recepção, tratamento e preparo desses dados implica que sua disponibilidade não ocorre propriamente em tempo real. Há um atraso entre a situação na via e a chegada da informação aos sistemas, os quais também demandam tempo de processamento e apresentação das informações atualizadas. Daí resulta que essas informações se refiram sempre a um passado próximo e não à situação presente. Assim, torna-se necessária a criação de ferramentas que possibilitem estimar a situação de momento, com base nos dados recebidos.

A estimação de tráfego consiste em atualizar, para o momento presente, as informações de tráfego defasadas - recebidas de sensores e outras fontes - tendo em vista os tempos de transmissão, tratamento e processamento dos dados. Assim, a estimação pode ser feita olhando-se para um intervalo à frente ao da coleta dos dados.

Nesta tese busca-se determinar métodos adequados para o processo de estimação e de seu uso na previsão de tráfego de curto prazo. A metodologia desta tese está baseada na análise comparativa de métodos de estimação da situação de tráfego e do uso desses métodos na previsão do tráfego de curto prazo.

A metodologia de estimação, proposta nesta tese, é apresentada no capítulo 3, e prevê a obtenção e organização dos dados de tráfego, assim como seu uso em modelos de estimação que permitem caracterizar a atual situação do tráfego (Figura 1-1).

Figura 1-1 – Fluxograma da metodologia proposta



Fonte: o autor

A obtenção inicial dos dados, e seu tratamento, não se mostrou trivial. Mas, foi necessária para que fosse possível sua utilização nesta tese, e em aplicações futuras. Para tal, foi necessária a criação de scripts específicos para a conclusão de diversos passos desta tese, o que se deu através da utilização das linguagens Python e R.

O pré-tratamento dos dados envolve a complementação dos dados, com base nos dados estimados - pelo próprio método de estimação - em intervalos anteriores. Também nessa etapa ocorre o tratamento de dados atípicos, através da aplicação de métodos que verificam o desvio da informação recebida, em relação ao padrão de dados do momento.

Os métodos de estimação avaliados, incluíram os modelos Espaço Temporais Autorregressivos com Média Móvel (STARIMA) e os modelos de Regressão para Dados em Painel, com e sem tratamento hierárquico, cujos resultados obtidos são apresentados no capítulo 5.

1.1 MOTIVAÇÃO

Durante 12 anos coordenando diversas áreas da Central de Operações da Companhia de Engenharia de Tráfego de São Paulo (CET-SP), o autor se deparou com inúmeros desafios relativos à gestão da informação em tempo real, que o levaram a pesquisar formas e ferramentas que pudessem otimizar os processos e parametrizar as decisões.

Dentre as possíveis fontes de dados se apresentam os sensores do tipo radar instalados nas vias do Município de São Paulo, cujos dados encontravam-se fechados. Até a presente data sua utilização é para sistemas de fiscalização do cumprimento de regras de trânsito e não para o controle dos fluxos de tráfego ou outros usos.

Durante o início desta tese foi obtido acesso aos dados dos sensores (radares) junto à Secretaria Municipal de Mobilidade e Transportes do Município de São Paulo (SMMT-SP), possibilitando a oportunidade de transformação desses dados em informações com valor agregado para a gestão dos fluxos de tráfego. Além da integração com outros sistemas, vislumbrando inclusive um uso futuro em sistemas dinâmicos de previsão de tráfego e de fornecimento de informação de tráfego em tempo real a outros sistemas de controle e gestão.

1.2 OBJETIVO

A pesquisa tem como objetivo definir uma metodologia para estimar as características espaço-temporais do tráfego, em condições dinâmicas (tempo próximo ao real), visando a previsão de tráfego de curto prazo, para uso em sistemas de controle e informação de tráfego.

A estimação de tráfego consiste em atualizar para o momento presente, as informações de tráfego defasadas, recebidas de sensores e outras fontes. A estimação das características de tráfego é necessária aos processos dinâmicos de previsão de tráfego e útil aos sistemas de controle de tráfego e de gestão operacional, e na implementação de cidades inteligentes, às quais o tráfego e o transporte se conectam através do ITS.

São utilizados dados de sensores instalados nas vias, de forma a minorar a dependência da disponibilidade de dados difíceis de obter (como matrizes de origem-destino dinâmicas, necessárias em abordagens mais detalhadas).

Como objetivos complementares temos:

- a) incorporação do aspecto espacial a fim de possibilitar a aplicação do modelo em redes de tráfego urbano;
- b) análise comparativa do uso de métodos de reconhecimento de padrão na estimação de tráfego para uso em ambiente urbano;
- c) avaliação do uso do método de estimação como ferramenta de previsão de tráfego de curto prazo.

1.3 QUESTÕES DE PESQUISA

- a) como é possível estimar as condições de tráfego no ambiente urbano?
- b) o uso de séries temporais é adequado para modelos de previsão de tráfego?
- c) quais características espaço temporais devem ser utilizadas nos modelos de estimação de tráfego de curto prazo?
- d) é possível melhorar o modelo de matriz de pesos espaciais utilizado nos sistemas de estimação – como por exemplo o modelo STARIMA?

1.4 ORGANIZAÇÃO

O presente texto está estruturado em 5 partes, iniciado pela Introdução, na qual foi apresentada a motivação e os objetivos da pesquisa. A segunda parte trata da vasta base teórica relacionada ao tema, incluindo os conceitos envolvidos na pesquisa. Na terceira parte é apresentada a metodologia proposta e na parte seguinte os estudos desenvolvidos com base

nos dados disponíveis. Por fim são apresentas as considerações finais e as propostas para trabalhos futuros.

Os Apêndices trazem a relação de artigos científicos relacionados à tese, relação de Trabalhos de Final de Curso e Iniciações Científicas relacionadas à tese e que foram coorientados pelo autor; uma aplicação prática do modelo de Dados em Painel para complemento de dados históricos de volumetria na Cidade de São Paulo; descrição do aplicativo STARIMA (PAZ; TRINTINALIA, 2021), os Scripts utilizados durante a tese; e por fim, informações sobre detecção de incidentes levantadas durante o estudo e que, por estarem fora do escopo, foram retiradas do texto principal.

2 REVISÃO DA LITERATURA

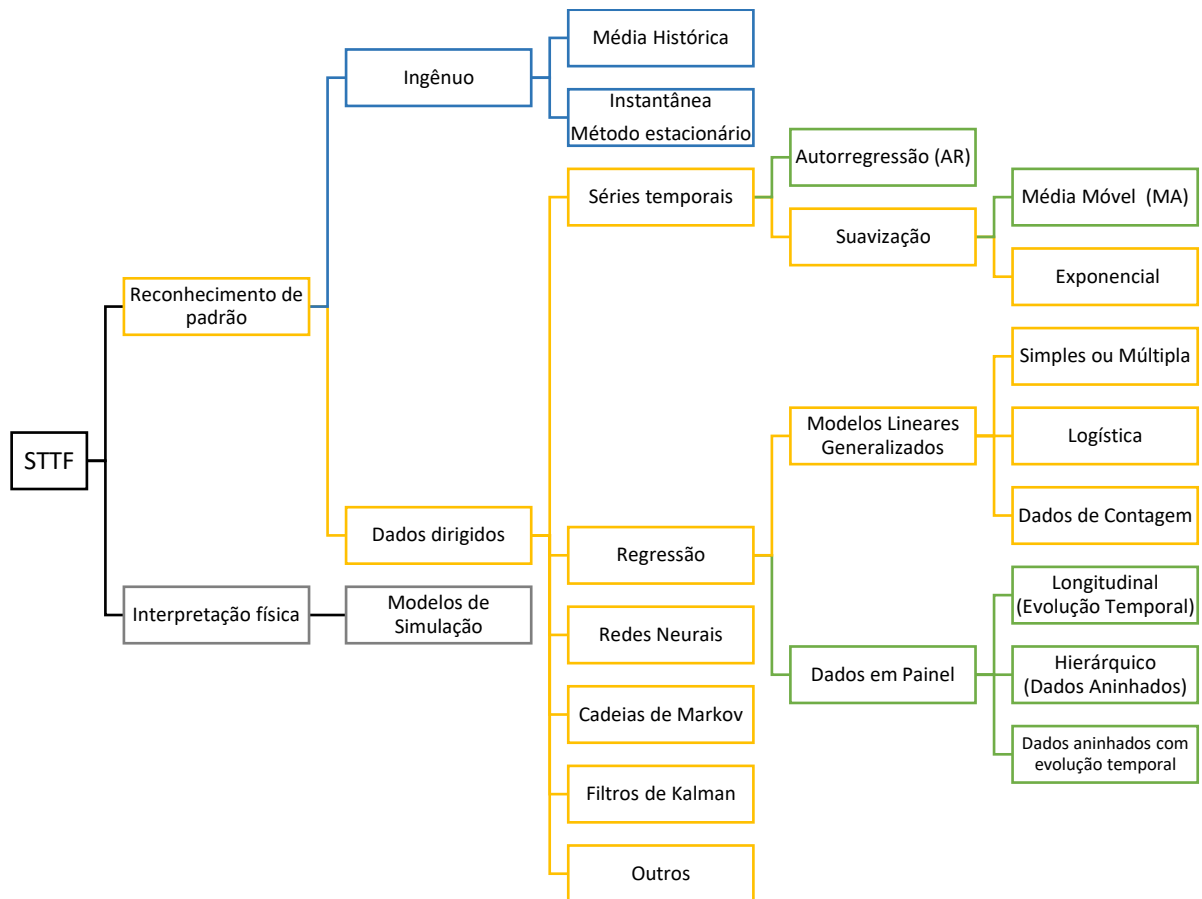
Neste capítulo é apresentada a revisão da literatura relacionada à estimação e previsão de tráfego, visando respaldar o modelo de estimação apresentado no capítulo seguinte.

2.1 PREVISÃO DE TRÁFEGO

Os sistemas de estimação e previsão de tráfego podem ser classificados com base no seu horizonte de previsão, sendo que os de curto prazo trabalham com horizontes de menos de um minuto até poucas horas à frente do momento do processamento e os de longo prazo chegam a atingir o horizonte de vários dias. A previsão de curto prazo das condições do tráfego feita em tempo próximo ao real, mostra-se viável como ferramenta de gestão integrada a diversas outras ferramentas de ITS, podendo ser utilizada nos Centros de Gerenciamento Operacional na tomada de decisão em tempo próximo ao real, comparando cenários de alternativas operacionais para situações em andamento. (BARROS, 2014; SLOBODEN et al., 2012; VLAHOGIANNI; KARLAFTIS; GOLIAS, 2014).

A Previsão de Tráfego de curto prazo pode ser classificada em dois tipos de abordagem: Reconhecimento de Padrão e Interpretação Física do Fenômeno de Tráfego (GENTILE; MESCHINI, 2011). De modo semelhante os sistemas de previsão podem ser divididos em 3 grupos, sendo os dois primeiros uma subdivisão do grupo Reconhecimento de Padrão: Previsões Ingênuas, Previsões baseadas em dados dirigidos e Previsões Baseadas em Modelos (VAN LINT; HOOGENDOORN; VAN ZUYLEN, 2005). A Figura 2-1 apresenta um resumo dos métodos utilizados nos modelos de STTF.

Figura 2-1 – Métodos de Previsão de Tráfego de Curto Prazo (STTF)



Obs. Os itens destacados em verde são os utilizados nesta tese nos processos de estimação de tráfego e serão descritos com maior detalhe nos itens seguintes. Os modelos AR e MA são a base do modelo STARIMA utilizado.
Fonte: do autor

Assim temos:

- a) reconhecimento de padrão baseada em inferência estatística de dados históricos: podendo ser caracterizados como modelos ingênuos, como a média histórica e os modelos instantâneos, e modelos com dados dirigidos, como as séries temporais, regressões (Lineares e com Dados em Painel), redes neurais, Cadeias de Markov e Filtros de Kalman, dentre outros. Busca identificar o padrão de tráfego de um dado ponto ou link da rede, e com base na projeção de novas demandas, ou de dados recentes do tráfego, estima a evolução futura.

São sistemas que permitem a previsão do tráfego sem a necessidade de construção da rede viária e calibração das características do fluxo, possibilitando assim menor esforço preparatório do modelo e frequentemente menor custo computacional para obtenção dos resultados. Apesar disto, informações e características da rede

viária podem ser incorporadas aos modelos através da inclusão de variáveis com características físicas ou operacionais, informações espaciais e de relacionamento.

Este tipo de abordagem é menos capaz de tratar situações inesperadas e não previstas e eventos atípicos, como acidentes, uma vez que os mesmos ocorrem de modo aleatório na rede e não são previamente registrados em quantidade suficiente para que o modelo possa se adaptar (GENTILE; MESCHINI, 2011; SEO et al., 2017). Criar e treinar esse tipo de sistema para as situações atípicas significaria uma infundável sequência de suposições e treinamentos de modelos e ainda assim seria possível a ocorrência de situações não definidas e treinadas previamente.

- b) interpretação física do fenômeno de tráfego: baseada na modelagem da interação entre demanda de viagem e a rede viária através da alocação do tráfego. Normalmente efetuada com o uso de simuladores de tráfego, que simulam a propagação do fluxo na rede, inclusive para os locais não monitorados em tempo real (PORTUGAL, 2005).

Nesta abordagem é possível efetuar previsões do tráfego também em situações não ocorridas anteriormente, efetuando ajustes na capacidade (oferta), nos casos de acidentes, obras na via, inclusão de novas vias e de manifestações populares entre outras.

Ajustes também podem ser feitos na demanda, em decorrência de eventos na via ou em locais de concentração de público ou da criação/fechamento de empreendimentos geradores de tráfego (shopping-centers, condomínios residenciais e comerciais, etc.) efetuando a redistribuição do fluxo pela rede com base nessas premissas. (GENTILE; MESCHINI, 2011; SEO et al., 2017).

Necessita de grande quantidade de dados dinâmicos difíceis de se obter, como a matriz origem destino (OD) dinâmica, e de calibração detalhada de diversas características do modelo. Os esquemas para incorporação de dados dinâmicos na abordagem baseada em modelos de tráfego são razoavelmente complexos e computacionalmente custosos (WANG; PAPAGEORGIOU; MESSMER, 2006).

Certamente, propostas de soluções híbridas tem surgido ao longo do tempo, mostrando-se promissoras ao utilizar as melhores características de cada solução (ALLSTRÖM et al., 2016; GLIŠOVIĆ et al., 2016; TSELENTIS; VLAHOGIANNI; KARLAFTIS, 2015; VAN LINT; HOOGENDOORN; VAN ZUYLEN, 2005).

2.1.1 Estimação do estado do tráfego ou previsão de tráfego?

Por vezes a expressão Estimação do Estado do Tráfego tem sido utilizada como sinônimo de Previsão de Tráfego em momento futuro, ao “estimar” a situação futura. Neste trabalho os dois conceitos se distinguem, conforme sugerido por (VAN LINT; HOOGENDOORN; VAN ZUYLEN, 2005):

- a) a Estimação do Estado do Tráfego consiste em determinar a situação do tráfego em determinado local num determinado momento e está normalmente relacionada à determinação da situação presente do tráfego. Esta é a etapa na qual ocorre o tratamento dos dados vindos de diversas fontes com o objetivo de caracterizar o tráfego. São incluídos principalmente estudos com uso de sensores ITS e dados de posicionamento de celulares e de veículos;
- b) a Previsão de Tráfego consiste em determinar as condições em que o tráfego se apresentará após algum tempo, podendo ser de curto ou longo prazo, conforme já descrito. Os estudos de Previsão de Tráfego são baseados nas situações presente e passada do tráfego, obtidas nos estudos de Estimação do Estado do Tráfego.

A estimação e os estudos de previsão estão relacionados às características do tráfego. Na literatura acadêmica foi possível verificar estudos relativos às principais características:

- a) velocidade e tempo de percurso (CHEN; CHIEN, 2001; GENTILI; MIRCHANDANI, 2018; HIRIOTAPPA et al., 2017; WANG; PAPAGEORGIOU; MESSMER, 2006)
- b) densidade (MUIIOZ et al., 2003; QIU et al., 2010; T. DARWISH; K. ABU BAKAR, 2015; WADDOO et al., 2016)
- c) fluxo de tráfego (CASTILLO et al., 2015; GIACOMO COMO, 2017; STATHOPOULOS; KARLAFTIS, 2003)
- d) nível de serviço (SAHA et al., 2016)

Boa parte dos estudos de estimação e previsão verificados partem do princípio de que os dados de Estimação estão disponíveis e que não existe a necessidade de tratar lacunas ou falhas. Em alguns casos é indicada a necessidade de uma fase inicial de tratamento dos dados para preparação, todavia não fazem menção da metodologia a ser empregada no tratamento de falhas. Muitos desses estudos, especialmente aqueles baseados em Reconhecimento de Padrão, trabalham apenas com a situação média e desconsideram as situações atípicas.

Conforme apontado em estudos anteriores, existe a necessidade de tratar essas lacunas e falhas nos dados para possibilitar o correto funcionamento do modelo de previsão (LI et al., 2014; LI; LI; LI, 2013; SERNEELS; VERDONCK, 2008; WANG; MAO, 2019).

Na mesma direção, os modelos de estimação e previsão podem incluir o impacto de fatores externos como condições meteorológicas (GONÇALVES, 2018; HOU et al., 2013) e incidentes (BURGHOUT; KOUTSOPOULOS; ANDREASSON, 2010; DIA; THOMAS, 2011; FHWA OFFICE OF HIGHWAY INFORMATION MANAGEMENT, 1998; WANG; QI; JIANG, 2018).

2.2 ESTIMAÇÃO E PREVISÃO DE TRÁFEGO ESPAÇO-TEMPORAL

Um dos desafios presentes nos sistemas de estimação e previsão de tráfego é a inclusão do aspecto espacial da rede de tráfego nos sistemas de previsão de curto prazo produzindo assim sistemas que tratem simultaneamente os aspectos espaço-temporais. (VLAHOIANNI; KARLAFTIS; GOLIAS, 2014).

Incorporar a característica espaço-temporal (ST) num modelo de Previsão de Tráfego de Curto Prazo (*Short Time Traffic Forecast – STTF*) pode propiciar um maior entendimento da evolução do tráfego na rede, uma vez que possibilita estimar como a mudança de padrões de tráfego, em determinado local, se propaga para o resto da rede viária.

Para a inclusão do fator temporal, os modelos de estimação e previsão de curto prazo são organizados e resumidos por intervalos, que podem variar de poucos segundos até alguns minutos, construindo-se séries temporais das características do tráfego. Dentre os modelos utilizados em análises temporais temos os modelos Autorregressivos Integrados de Média Móvel (*Autorregressive Integrated Moving Average - ARIMA*) que é um modelo univariado (LAÑA et al., 2018) e o tratamento através de Modelos de Regressão para Dados em Painel Longitudinal ou com medidas repetidas (FÁVERO; BELFIORE, 2017).

A incorporação do aspecto espacial, se dá através da análise simultânea de diversos locais de estudo. Essa inclusão pode ser feita com base em matrizes de peso, como no STARIMA, ou através da inclusão de variáveis explicativas, no caso de análises multivariadas.

Em revisão científica sobre STTF espaço temporal (ST-STTF), foram verificadas 130 publicações entre os anos de 1984 e 2016, sendo a maioria delas apresentadas após 2008. Contatou-se que 45% dos estudos utilizaram modelos de Aprendizado de Máquina, 37% Modelos Estatísticos e 18% utilizaram modelos mistos com as duas abordagens. Verificou-se

que, embora os modelos de Aprendizado de Máquina tenham ganhado força no início da década, não existe garantia de que sejam melhores do que os modelos estatísticos avançados baseados no modelo autorregressivo STARIMA (ERMAGUN; LEVINSON, 2018).

Noutra revisão que incluiu 211 publicações do período de 1984 a 2018 constatou-se que nos últimos anos os métodos estatísticos paramétricos e não paramétricos aumentaram sua participação nos estudos, enquanto os métodos baseados em Redes Neurais apresentaram certa redução (PAVLYUK, 2019).

Destas revisões, constata-se que este é um tema ainda aberto à pesquisa, motivo pelo qual foi escolhido para esta tese. Assim, optamos por comparar dois métodos, o STARIMA, que tem abordagem autorregressiva e univariada, e a abordagem por Dados em Painel longitudinais e hierarquizados, que pode ser multivariada e utiliza agrupamentos, se aproximando dos modelos de aprendizado de máquina.

2.2.1 Estimação com uso de modelos autorregressivos espaço-temporais

O modelo STARIMA é um caso particular do modelo Vetor Autorregressivo com Média Móvel (VARMA), voltado para aplicações espaço temporais, que permite a redução significativa dos parâmetros a serem avaliados ao incorporar a topologia espacial através do uso da matriz de pesos, nas quais os locais mais próximos exercem maior influência do que aqueles mais distantes (Ver item 2.3). Na mesma direção, por ser um modelo autorregressivo, o passado recente exerce influência maior do que o mais distante na situação atual. (KAMARIANAKIS; PRASTACOS, 2003).

A primeira aplicação do STARIMA na ST-STTF foi apresentada por (KAMARIANAKIS, 2003). Estudos posteriores buscaram incluir a característica dinâmica do fluxo e se utilizaram da associação do modelo STARIMA com o modelo DTRP (Dynamic Turn Ratio Prediction), sendo necessária a definição de taxas de conversão de e para os links e a atualização em tempo real da matriz de pesos (DUAN et al., 2019; GUAN; HUANG; QU, 2018; MIN et al., 2009). Este tipo de aplicação dinâmica depende de grande quantidade de sensores monitorando a rede de modo a possibilitar o acompanhamento ou determinação das taxas de conversão.

Outra abordagem dinâmica para o modelo STARIMA efetuou a atribuição de pesos de acordo com a velocidade nos links de montante e jusante. Neste caso, a aplicação também

considera características do fluxo do tráfego e atualiza os pesos da matriz a cada passo (CHENG et al., 2011, 2014).

Embora uma das vantagens do STARIMA seja a definição de parâmetros que englobem todos os locais de estudo, modelos com enfoque local como o *Localised STARIMA* (LSTARIMA) (CHENG et al., 2014), posteriormente ajustado para o *Generalised STARIMA* (GSTARIMA) através de matrizes diagonais de peso dinâmico (MIN; HU; ZHANG, 2010), já foram avaliados. Nestes casos, utiliza-se o poder do STARIMA em se apropriar do impacto dos locais próximos sobre o ponto avaliado. No entanto, perde-se a vantagem do reduzido número de parâmetros a serem estabelecidos na construção do modelo, devido à necessidade de criar um modelo, com seus respectivos parâmetros, para cada local.

Além da área de previsão de tráfego, os estudos com aplicação do modelo STARIMA ocorrem em áreas como segurança pública, economia, hidrologia, disseminação de doenças, meteorologia e poluição do ar, o que mostra o potencial de sua aplicação.

2.2.2 Estimação com uso de regressão para dados em painel

Os modelos de regressão, para Dados em Painel, permitem a análise de dados com medidas repetidas em diversos instantes e são também conhecidos como modelos longitudinais. Para este tipo de estudo, as observações de cada indivíduo são organizadas em ordem cronológica, possibilitando a análise das séries temporais.

Os modelos de regressão multinível para dados em painel (*Generalized Linear Multlevel Model* – GLMM) possuem estrutura aninhada de dados, que permite efetuar análises de uma variável dependente, considerando a evolução dos indivíduos ao longo do tempo e as diferenças entre indivíduos, agrupados de acordo com suas características. Podem inclusive ocorrer agrupamentos dentro de agrupamentos, o que caracteriza os painéis multinível ou modelos hierárquicos. Os modelos hierárquicos podem ser lineares (*hierarchical linear models* HLM) como os utilizados nesta pesquisa para os dados de velocidade, e não lineares (*hierarchical non linear models* – HNM). Os modelos não lineares são destinados aos estudos com dados de contagem, com valores discretos e não negativos, como ocorre no caso dos modelos para dados de fluxo de tráfego (FÁVERO; BELFIORE, 2017).

O uso de modelos espaciais, baseados em dados em painel, foi anteriormente abordado em estudos econométricos (BALTAGI; BRESSON; PIROTTE, 2007; ELHORST, 2003, 2017) e

permite a análise do impacto espacial, tanto na variável dependente como nas variáveis explicativas e nos termos de erro (ANSELIN, 2021).

Na área de tráfego, o GLMM tem sido frequentemente utilizado nas análises de previsão de acidentes (Apêndice E). No entanto, seu uso nos modelos de estimação e previsão de tráfego é menos frequente.

Na área de transportes podem ser encontradas aplicações de modelo Longitudinal com efeitos fixos para escolha de rotas (BEN-ELIA; SHIFTAN, 2010), análise do fluxo de ciclistas (TANG; THAKURIAH, 2012) e da demanda de passageiros no transporte coletivo (VIEGO; VOLONTÉ, 2017). Nestes casos, considera-se a evolução temporal e outras variáveis como condições meteorológicas, valor do combustível e tarifa do transporte público. Noutra abordagem Longitudinal com efeitos fixos, a previsão da velocidade foi efetuada apenas utilizando os dados da própria série temporal local (GEORGESCU; ZEITLER; STANDRIDGE, 2012). A abordagem longitudinal com efeitos fixos e aleatórios foi empregada em análise da quantidade de quilómetros rodados mensalmente com base em variáveis sócio-econômicas e também na análise do impacto de obras na via sobre a velocidade do tráfego (KIM, 2022).

A estrutura hierárquica espacial, baseada na relação entre os locais, foi utilizada em algumas pesquisas de determinação da velocidade média ou velocidade operacional (V85 – velocidade praticada pelo 85º percentil dos usuários da via). A hierarquização espacial pode receber estruturas diferentes, conforme o caso de estudo, separando-se em níveis: a direção do fluxo, a via e os trechos de cada via (BASSANI et al., 2016a; TREMBLAY; CIRILLO; BASSANI, 2021); a via, as seções da via e as faixas de rolamento (BASSANI et al., 2014, 2016b) ou ainda, num primeiro nível, os dados podem estar agrupados pelas características do fluxo, no segundo para os tipos de via e no terceiro para o tipo de região onde a via se insere (ISLAM; EL-BASYOUNY, 2015).

A hierarquização permite inclusive a aplicação de valores estimados para locais conhecidos servirem de insumo para locais onde não se tenham dados históricos disponíveis, com base na semelhança entre os locais agrupados, levando em conta atributos físicos (curva, declividade, etc.) e operacionais (tipo de fluxo, presença de pedestres, etc.) (TREMBLAY; CIRILLO; BASSANI, 2021).

Observa-se que os modelos de Dados em Painel espaço-temporais, com pequena dimensão espacial, podem utilizar matrizes de covariância espacial, além de poderem ser descritos como um modelo de Vetor Autorregressivo (VAR) (GRIFFITH, 2010).

Como se verifica, os modelos com Dados em Painel utilizados na área de previsão de tráfego ou se utilizam do modelo longitudinal, ou utilizam a estrutura hierárquica sem considerar especificamente a série temporal dos dados. Assim, nossa pesquisa inova, ao aplicar o modelo hierárquico considerando no primeiro nível a informação da série temporal da variável dependente.

2.2.3 Estimação com uso de modelos de simulação

Os modelos de Simulação de Tráfego, por sua vez, atuam como ferramentas de previsão de tráfego de curto e longo prazo e são inerentemente dotados de análise espaço-temporal uma vez que usam como interface a montagem de modelos da rede viária, por onde os usuários vão se deslocar, e sua aplicação ocorre através da alocação do tráfego no modelo, de modo distribuído ao longo de determinado período. Esses modelos incorporam as interações espaço temporais entre os usuários da via.

O uso dos simuladores de tráfego como ferramenta de previsão de tráfego de curto prazo se dá através da utilização de Alocação Dinâmica do Tráfego (Dynamic Traffic Assignment – DTA) e foi apresentado em diversos textos científicos (BEN-AKIVA et al., 2002; KUCHARSKI; KOSTIC; GENTILE, 2017; XU; KONG; LIU, 2012).

A DTA consiste na atualização constante do modelo com os dados que representem a situação (estado) do tráfego na rede (SLOBODEN et al., 2012). Assim, a alocação dinâmica (em tempo próximo ao real) depende da utilização de dados atualizados das características do tráfego, os quais são obtidos através de metodologias de estimação do tráfego como as que propomos nesta tese.

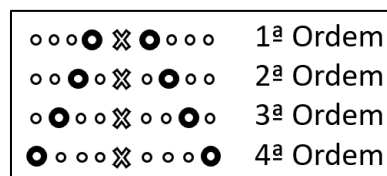
2.3 MATRIZ DE PESO ESPACIAL

A “Lei de Tobler” também conhecida como Primeira Lei da Geografia diz que “Todas as coisas estão relacionadas com todas as outras, mas coisas próximas estão mais relacionadas do que coisas distantes” (TOBLER, 1970). Dela verifica-se a necessidade e possibilidade de tratar análises espaço-temporais considerando as características de “relacionamento por proximidade” entre os locais de estudo.

Isto possibilita a abordagem utilizada por Cliff & Ord (1975) no modelo Space–Time Autoregressive Integrated Moving Average Regressive (STARIMAR) e por Pfeifer & Deutsch (1980a) no desenvolvimento do STARIMA. Eles utilizam as matrizes de peso para incluir os aspectos espaciais. O mesmo tipo de matriz é utilizado também em outras abordagens como k Nearest Neighbor (KNN) (CAI et al., 2016), Redes Neurais (FENG et al., 2021) e em estudos de definição da velocidade em trechos de rodovia, com uso de Regressão Linear Multivariada (WANG; GUO; TARKO, 2020). Esse tipo de matriz é também empregado nos modelos de Dados em Painel em outras áreas de estudo (ANSELIN, 2021; ELHORST, 2003; ELHORST; HALLECK VEGA, 2017; ZHOU; WANG, 2018).

Em seus estudos de desenvolvimento e aplicação do modelo STARIMA, Pfeifer & Deutsch (1980a) estabeleceram matrizes espaciais de acordo com a vizinhança por contiguidade (VC), dividindo-se os vizinhos de acordo com a “ordem” de vizinhança. Essa ordem pode ser unidirecional ou bidirecional. A vizinhança bidirecional é utilizada nos estudos da geografia. Nos estudos de tráfego, onde se considera o sentido do fluxo do tráfego nas vias, é possível o modelo unidirecional que é o mais simples (Figura 2-2).

Figura 2-2 – Ordem de vizinhanças unidirecionais utilizadas em matrizes espaciais



Fonte: adaptado de (PFEIFER; DEUTSCH, 1980a)

A aplicação das matrizes de peso no modelo STARIMA possibilita uma clara visão das duas dimensões “impacto” e “relacionamento” às quais podemos chamar respectivamente de “peso” e “vizinhança”. Os valores de cada termo w_{nm}^l da matriz W^l representam o peso (intensidade/impacto) atribuído à relação entre os locais “nm” e “nj” em determinada ordem “l” de vizinhança. Os pares “nm nj” que não pertencem à ordem “l” recebem o valor 0. De modo geral, pode-se construir no máximo $N - 1$ matrizes de peso espacial de dimensão $N \times N$, sendo “N” o número de locais com dados disponíveis.

2.3.1 Pesos da matriz na ST-STTF

Uma vez que os pesos atribuídos aos termos $w_{nm}^l n_j$ representam a intensidade de impacto de um ponto, trecho ou região “nm” em outro “nj”, a escolha do correto parâmetro de medição dessa intensidade para cada estudo é fundamental ao sucesso de aplicação das matrizes de peso. Nos estudos com modelos STARIMA verifica-se a aplicação dos tipos de peso indicados na Tabela 2-1.

Tabela 2-1- Pesos e Vizinhanças em ST-STTF com STARIMA

Vizinhança	Pesos	Referência
Assimétrica por Contiguidade	Pesos Unitários Normalizados	(KAMARIANAKIS; PRASTACOS, 2003) (KAMARIANAKIS; PRASTACOS, 2005)
	Pesos Unitários Normalizados e fluxo de montante dividido pelo fluxo do local avaliado, proporcional ao tempo de percurso	(TAO; SUN; BOUKERCHE, 2019)
	Correlação de Pearson	(LIN et al., 2009)
Simétrica por Contiguidade	Pesos Unitários Normalizados	(MIN; HU; ZHANG, 2010)
	Pesos Unitários	(WANG; TSAPAKIS; ZHONG, 2016) (PAVLYUK, 2017)
	Distância com pesos por tipo de relacionamento	(WU et al., 2015)
	Taxa de conversão de i para j – dinâmico	(GUAN; HUANG; QU, 2018) (MIN et al., 2009)
Simétrica por Lag de Percurso	Pesos Unitários	(DING et al., 2011) (DUAN et al., 2016)
	Taxa de conversão de i para j – dinâmico	(DUAN et al., 2019)
	Coeficiente de determinação entre os trechos	(DIAMANTOPOULOS et al., 2013) (SALAMANIS et al., 2016) (SALAMANIS et al., 2015)
Assimétrica por Lag de Percurso	Diferença da velocidade dividida pela distância - dinâmico	(CHENG et al., 2011)
	Diferença da velocidade de montante pela de jusante - dinâmico	(CHENG et al., 2014)
	Pesos Unitários Assimétricos	(MIN; WYNTER, 2011) (DARAGHMI; YI; CHIANG, 2014)

Fonte: do autor

Como pode se verificar, na Tabela 2-1, temos diversos tipos de peso, sendo que vários modelos se utilizam de pesos unitários, onde não se distingue a intensidade de relação entre pontos mais distantes, ou menos correlacionados, em relação àqueles mais próximos ou mais correlacionados. Parte dos estudos que usam pesos unitários efetuam a normalização das matrizes, situação na qual as linhas das matrizes são tratadas para gerar a soma igual a 1. Isto

simplifica as operações matriciais, otimizando o processo. Todavia, segundo Anselin (2021), o tema da normalização das matrizes de peso é algo que ainda tem suscitado diversas pesquisas, havendo os que defendem e os que condenam sua utilização.

Pesos baseados na distância entre os pontos de medição, são diretamente inspirados na Lei de Tobler. Pesos que utilizam as características do tráfego também têm sido utilizados por alguns pesquisadores, como nos casos em que se considera a velocidade e o fluxo. Isto aproxima o modelo à teoria de fluxo do tráfego, com potencial aumento na acurácia. Na mesma direção, modelos que utilizam taxa de conversão, buscam incorporar as variações do fluxo em cada trecho de estudo, considerando as porcentagens de veículos que seguem na via e aqueles que efetuam conversões nos cruzamentos.

Matrizes de peso simétricas são as mais comuns nos estudos das diversas áreas da ciência, e consideram que o impacto de um local “nm” sobre o local “nj” é o mesmo de “nj” sobre “nm”. Todavia, na área de tráfego, a situação de fluxo num ponto/trecho de via impacta de modo diferente o fluxo no ponto/trecho a montante e a jusante. Assim, os modelos assimétricos, que consideram a diferença entre os fluxos de montante e jusante, devem ser considerados.

Por fim, um dos grandes desafios de modelos espaço temporal autorregressivos (STAR) relaciona-se à dinâmica do fenômeno estudado, como no caso da ST-STTF, onde as condições de fluxo de tráfego se alteram constantemente (PAVLYUK, 2020). Nesse sentido podem ser consideradas matrizes dinâmicas, que de um lado incorporam as constantes alterações do fluxo, mas que de outro lado tornam o modelo mais complexo.

Nesta tese comparamos modelos estáticos com diversos tipos de peso, a fim de melhor entender a importância desse aspecto e de verificar quais variáveis de relacionamento trariam maior ganho à estimação.

2.3.2 Vizinhaças da matriz na ST-STTF

As matrizes de peso espaciais podem ser tratadas separadamente de acordo com a ordem de vizinhança ou de modo unificado/agrupado. No uso separado, conforme proposto por (PFEIFER; DEUTSCH, 1980b, 1980a) cada matriz W^l , inclui apenas os valores de relacionamento (pesos) entre os elementos de estudo de “l-ésima” ordem. No uso agrupado

(AG) todas as vizinhanças consideradas são incluídas numa única matriz de ordem 1 (PFEIFER; BODILY, 1990; RAO; ANTUNES, 2004).

A definição das ordens de vizinhança de matrizes separadas depende do fenômeno estudado e da distribuição dos locais avaliados, sendo determinada durante o processo de calibração do modelo. Nos estudos de tráfego o modelo de vizinhança mais utilizado é o unidirecional. Todavia, não se descarta a utilização de modelos bidimensionais onde o impacto do fluxo na rede viária é considerado independentemente dos trajetos efetuados (ERMAGUN; LEVINSON, 2019; GUAN; HUANG; QU, 2018).

As matrizes espaciais do tipo Vizinhança por Contiguidade (VC) propostas por (PFEIFER; DEUTSCH, 1980b, 1980a) são definidas unicamente pela localização geográfica, onde o vizinho é o primeiro local verificado numa direção, de modo que, para cada “Ordem espacial” é criada uma matriz de pesos separada. Na matriz de Ordem 1 são incluídos os primeiros vizinhos, e os vizinhos destes (ou segundos vizinhos) são inseridos na matriz de Ordem 2 e assim por diante.

O outro tipo de vizinhança presente nos estudos de ST-STTF é definido pelo *Lag* de percurso (LP) entre dois pontos de estudo. A forma de determinação do LP está detalhada no item 3.3. Neste caso, na primeira ordem são incluídos os locais percorridos em até um LP, na segunda ordem os que são percorridos em 2*LP etc (CHENG et al., 2014; MIN; WYNTER, 2011). Este tipo de vizinhança incorpora uma característica importante do tráfego de veículos, pois incorpora o tempo que um grupo de veículos que parte de um local chega a outro.

Numa abordagem diferente a ordem de vizinhança também utiliza o *Lag* temporal para definir a ordem de vizinhança, todavia são incluídos na matriz apenas os 10 pares de locais (geograficamente vizinhos ou não) com maior Coeficiente de Determinação - CoD (DIAMANTOPOULOS et al., 2013; SALAMANIS et al., 2016). Esta abordagem implica em grande esforço na análise de correlação entre as séries temporais de todos os N locais de estudo. Devem ser consideradas as muitas combinações possíveis entre todos os pontos e suas séries defasadas, de acordo com os diversos k intervalos a serem avaliados.

No modo AG, todos os pesos w_{nm}^l são agrupados numa única matriz W , como se todos os pares “ nm n_j ” tivessem a mesma ordem de vizinhança (RAO; ANTUNES, 2004), sendo mais simples e fácil de identificar e estimar os parâmetros da função (PFEIFER; BODILY, 1990). Na mesma direção, a aplicação de pesos obtidos do inverso da distância possibilita o uso de matrizes agrupadas numa única ordem (BIZ, 2014).

Embora o modelo agrupado já tenha sido aplicado em outras áreas de estudo (PFEIFER; BODILY, 1990; RAO; ANTUNES, 2004), durante a pesquisa bibliográfica não foi identificada sua aplicação nos modelos STARIMA de ST-STTF. Por esse motivo, na etapa de análise das matrizes, será apresentada a comparação da matriz agrupada (AG) com os modelos de Vizinhança por Contiguidade (VC) e por *Lag* de Percurso (LP).

Por fim, verificou-se a utilização de modelos com vizinhança assimétrica na ST-STTF, havendo os que consideraram apenas a relação de vizinhança do link de montante, de forma que o link à jusante não impacta no ponto de conexão (KAMARIANAKIS; PRASTACOS, 2005; LIN et al., 2009). E, aqueles que consideram as condições de fluxo à montante e à jusante, sendo atribuído peso para o trecho à jusante apenas no caso deste apresentar tráfego congestionado (MIN; WYNTER, 2011).

Os estudos mais recentes com STARIMA têm utilizado modelos que consideram a relação de vizinhança simétrica com bons resultados. Como será discutido no item 3.3, consideramos que, no caso de ST-STTF, tanto o impacto causado pelo ponto anterior (montante) como posterior (jusante) devem ser incluídos no modelo, pois ambos causam impacto no local estudado.

2.4 MODELOS DE CONTROLE PREDITIVO

Diversos sistemas de controle, como os de semáforos inteligentes, estão baseados em Modelos de Controle Preditivo (MPC - Model (Based) Predictive Control). Segundo Bordons & Camacho (1962) o MPC “não é uma estratégia específica de controle e sim um grande grupo de métodos de controle que fazem uso explícito do processo para obter o sinal de controle minimizando a função objetiva”. O MPC segue uma estratégia cíclica de 3 passos (BORDONS; CAMACHO, 1962):

- a) A cada passo é feita uma previsão do valor futuro em determinado horizonte $t+n$, ou seja, instante t acrescido de n intervalos de tempo. O valor de n depende do sistema de controle que utilizará o dado. Esta previsão depende dos valores atuais e passados. Os valores previstos para os instantes $t+1$ a $t+n$, e conseqüentemente os sinais de controle que serão aplicados nesses instantes, também podem ser considerados como base de informação para o passo iterativo de otimização;

- b) O grupo de sinais futuros é calculado pela otimização de determinado critério que busca torná-lo próximo à trajetória de referência. Este critério busca minimizar a diferença entre o sinal de saída previsto e a trajetória de referência prevista. A solução é normalmente interativa e depende de definições sobre a estrutura de controle futuro;
- c) O sinal é atualizado a cada ciclo do processo e enviado ao sistema de controle.

Segundo Zhou et al. (2017), o uso do MPC em controladores semafóricos centralizados se mostra superior à aplicação de sistemas com controles descentralizados. Na mesma direção, são mais adequados os sistemas semafóricos centralizados que consideram a hierarquização da rede, criando-se agrupamentos de cruzamentos em sub-regiões, que são agrupadas em regiões e estas como uma rede completa. Neste caso, o modelo hierárquico pode ser aplicado com base nas relações entre os elementos da rede, podendo incluir vias paralelas ou trechos de corredores de tráfego. Na mesma direção, o MPC hierárquico pode ser aplicado na definição de defasagens semafóricas (YE et al., 2016).

O módulo de Previsão é parte fundamental do MPC, e deve ser capaz de assimilar as variações que ocorrem ao longo do tempo, efetuando a melhor previsão possível, já que os sinais que serão aplicados no sistema de controle estarão baseados nessa previsão. Também deve ser capaz de permitir que as previsões sejam calculadas em tempo real e ainda, ser intuitivo e permitir uma análise teórica (BORDONS; CAMACHO, 1962; RONCOLI; PAPAMICHAIL; PAPAGEORGIOU, 2016). Assim, verifica-se que qualquer método de estimação que atenda a esses pré-requisitos pode ser utilizado em aplicações de MPC.

2.5 DADOS DE TRÁFEGO

A caracterização da situação do tráfego pode ser baseada em parâmetros como, por exemplo, velocidade, fluxo e densidade. Dessa forma, para o desenvolvimento de estudos e na aplicação de modelos reais é necessária a obtenção de informações quantitativas desses parâmetros.

2.5.1 Fontes de dados dinâmicos de tráfego

Os dados utilizados nos sistemas de previsão de tráfego podem ser provenientes de diversas fontes, que incluem desde informações automáticas, geradas por sistemas ITS, até informações verbais fornecidas por usuários da via.

As tecnologias disponíveis como fontes de informação aos sistemas dinâmicos de estimação e previsão de tráfego podem ser agrupadas em 3 categorias, de acordo com a forma como são obtidos os dados, conforme proposto por Parkany e Xie (2005). As duas primeiras são também utilizadas de modo mais direto e a terceira pode ser utilizada como fonte complementar dos sistemas dinâmicos, ou em sistemas *off-line*:

- a) Dados da Via: dados provenientes de sensores fixos, instalados na via, como radares, contadores, laços detectores e câmeras que são automaticamente gerados e transmitidos às centrais operacionais. Possuem o inconveniente de nem sempre estarem posicionados de modo otimizado à detecção de incidentes e de sua instalação e manutenção serem complexas, muitas vezes envolvendo impacto ao fluxo do tráfego;
- b) Dados de Veículo-Sonda (*Probe-Vehicle*): dados de posicionamento, utilizando o Global Positioning System (GPS) e eventualmente outros tipos de sensores instalados nos veículos, assim como celulares utilizados pelos motoristas e passageiros. São transmitidos automaticamente às centrais de controle através de redes sem fio e permitem a obtenção de informações em diversos trechos da via, dependendo apenas da movimentação do veículo-sonda. Com o avanço da automação veicular e com o massivo uso de celulares a quantidade de informações desse tipo de fonte tem crescido. Tem por inconveniente a possível ausência de informações em períodos ou dias de menor fluxo, decorrente da menor disponibilidade de veículos-sonda e o fato de que os dados gerados por esse tipo de fonte não são públicos;
- c) Interface Humana: dados informados às centrais operacionais por motoristas, usuários da via, agentes públicos, moradores e imprensa ou obtidos por meio de monitoramento de câmeras de vídeo. Essa informação pode ser recebida através de ligações telefônicas, mensagens de texto, voz ou vídeo ou através de aplicativos de informação de incidentes.

Além das fontes já descritas, existem atualmente diversos aplicativos instalados em celulares que podem também servir de base para os sistemas de estimação e previsão de tráfego e ainda, de identificação de incidentes, sendo que em alguns casos eles podem ser diretamente integrados via interface de programação (PAVLYUK, 2019; WADDO et al., 2016).

2.5.2 Parâmetros de caracterização do tráfego

O tráfego pode ser caracterizado com base em parâmetros como velocidade pontual ou instantânea, velocidade média de percurso, fluxo ou volume horário, densidade, ocupação, tempo de percurso e atrasos em fila dentre outras.

Dentre diversas variáveis quantitativas que caracterizam o tráfego, através do uso de sensores dotados de áreas de detecção, como laços detectores magnéticos ou virtuais, é possível obter (MAERIVOET; DE MOOR, 2005; TRANSPORTATION RESEARCH BOARD, 2016):

- a) Velocidade pontual: de acordo com os tempos e momentos de ativação e desativação dos laços detectores;
- b) Brecha (*gap*) Temporal: tempo decorrido entre uma detecção e outra, considerando apenas o período em que os laços não estão acionados;
- c) Intervalo (*headway*) temporal: de acordo com o tempo decorrido entre acionamentos sucessivos dos laços de detecção;
- d) Ocupação: Tempo de acionamento de um laço detector por um veículo;
- e) Ocupação no intervalo: Soma dos tempos de acionamento do laço detector, ou seja, soma de todas as ocupações ocorridas num determinado intervalo de tempo;
- f) Densidade: no caso de sensores que trabalham com laços indutivos duplos, de modo generalizado, é possível obter a densidade pontual dividindo-se a ocupação no intervalo pela extensão dos laços de detecção;
- g) Fluxo: eventualmente também chamado de volume nos textos de engenharia de tráfego, corresponde à quantidade de veículos que passam por uma seção de via, durante determinado tempo.

Para todas as variáveis acima é possível e usual a obtenção de valores médios por intervalo de medição.

Com base num sistema dotado de vários sensores, capazes de identificar os veículos por alguma forma como a leitura automática de placa ou identificação por radiofrequência (RFID), é possível identificar o veículo no momento de sua passagem pelo sensor. Com a identificação de um veículo em dois pontos diferentes da rede é possível obter:

- a) Tempo de Percurso: tempo gasto para percorrer um trecho de via. Pode ser obtido com o apoio de sensores que identifiquem os veículos nos pontos de entrada e saída do trecho;
- b) Velocidade de Percurso: obtida pelo quociente entre tempo de percurso e a distância percorrida. Neste caso também é possível se obter a média dos valores.

2.5.3 Intervalos de agrupamento dos dados

A montagem das séries temporais utilizadas em estudos de estimação e previsão de tráfego inclui o agrupamento dos dados em intervalos de tempo pré-definidos. Assim, por exemplo, o fluxo pode ser obtido pela contagem de veículos que passam num sensor, num determinado intervalo de tempo. A velocidade pode ser definida como a média ou a mediana das velocidades desses veículos no mesmo intervalo.

No caso dos sistemas semafóricos, a atualização do sistema pode ser feita durante o ciclo, em passos de 1 ou mais segundos, ou em períodos de vários segundos ou minutos. A primeira opção depende de sensores fornecendo dados em tempo real, um grande poder de processamento, para o tratamento desses dados, e de comunicação constante entre os sensores e o sistema de controle centralizado. Nos demais casos, é possível efetuar a otimização de um sistema de previsão de demanda de tráfego, em tempo próximo ao real, em intervalos periódicos, pré-definidos, da ordem de minutos (RONCOLI; PAPAMICHAIL; PAPAGEORGIU, 2016). Assim, o sistema tem mais tempo para efetuar a estimação e enviar os dados aos demais sistemas de controle, podendo ser aplicado em situações com menor poder de processamento ou de comunicação limitada.

As previsões feitas para intervalos iguais ao do tempo de ciclo semafórico, permitem que todo o fluxo que passa durante o ciclo pode ser agrupado e caracterizado (LIN et al., 2012). A perda de precisão tende a ser menos significativa em redes de maior porte, viabilizando sua aplicação (ABOUDOLAS; PAPAGEORGIU; KOSMATOPOULOS, 2009).

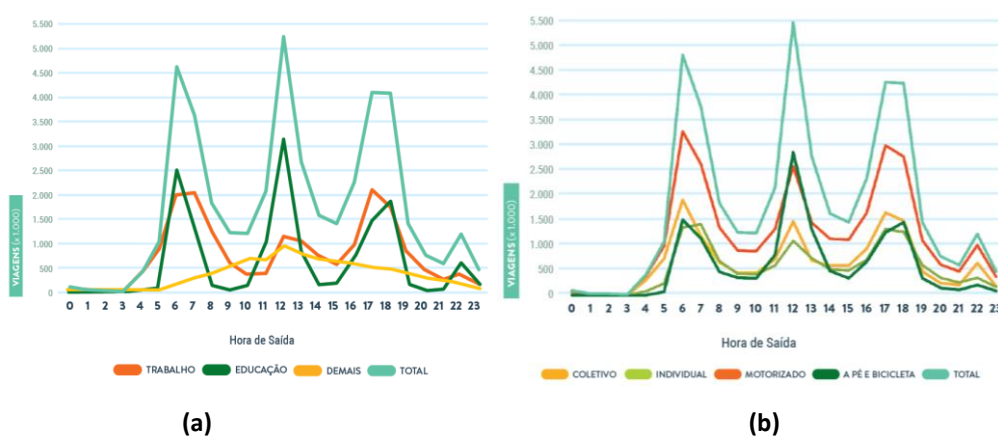
2.6 SAZONALIDADE

A sazonalidade do tráfego ocorre ao longo: dos meses do ano, com redução do fluxo nos meses de férias escolares; dos dias da semana, com redução da demanda aos finais de semana; e ao longo das horas do dia, com concentração da demanda nos horários de pico da

manhã e da tarde. O tema já foi alvo de diversos estudos de engenharia de tráfego como, por exemplo o realizado por Stathopoulos e Karlaftis (2001) e interessa aos sistemas de previsão de tráfego em tempo real (CHEN et al., 2012; YANG et al., 2016).

A sazonalidade diária é facilmente identificada em pesquisas OD (Figura 2-3). Também pode ser percebida através do acompanhamento do monitoramento dos congestionamentos na cidade ao longo dos dias e das semanas do ano (Figura 2-4), por exemplo através de rastreamento por GPS. Pode também ser diretamente verificada através de séries temporais de contagens no sistema viário.

Figura 2-3 - Flutuação horária das viagens por motivo (a) e por modo (b)

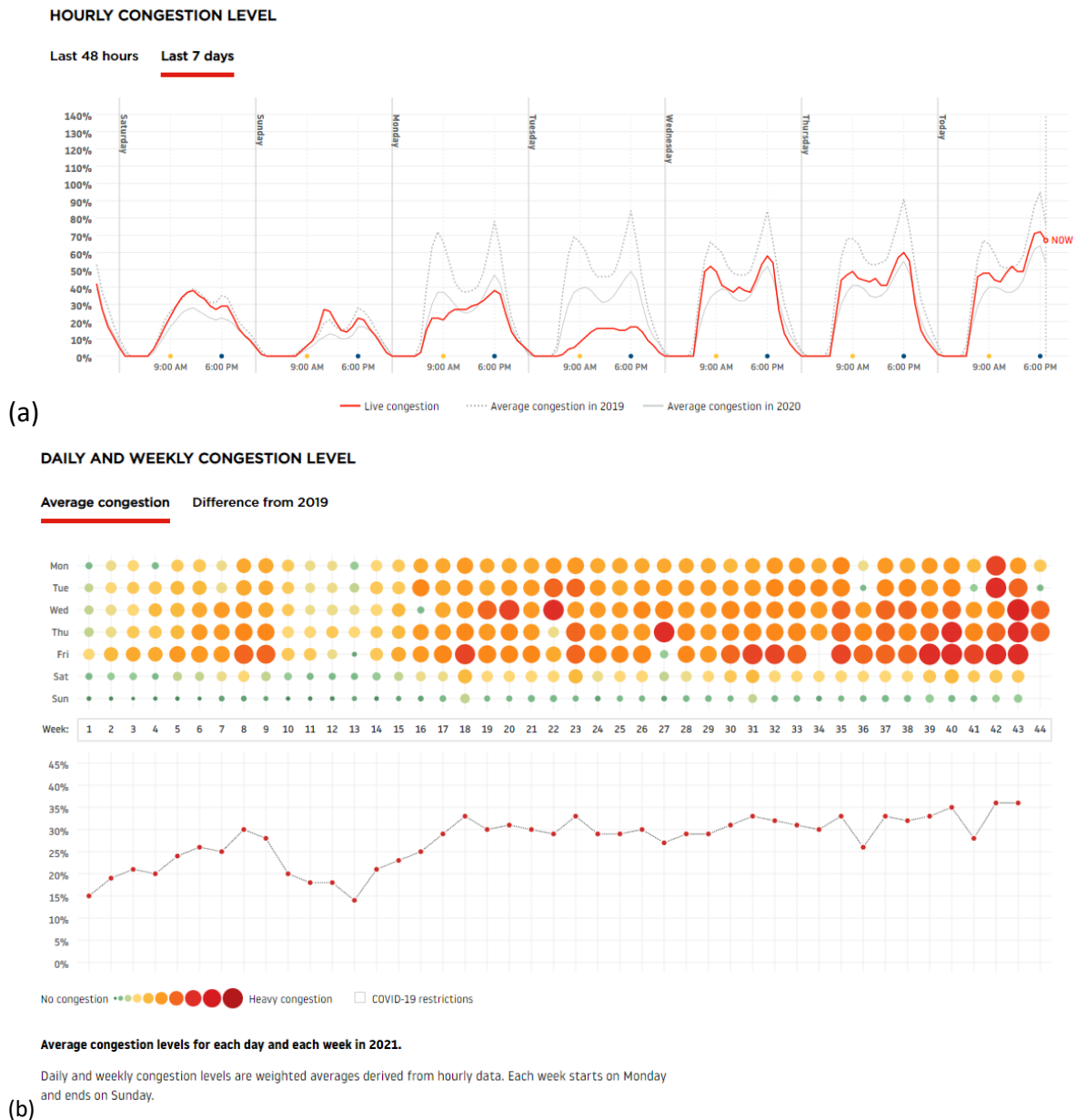


Fonte: (METRÔ SP, 2017)

A incorporação da sazonalidade nos modelos de estimação ou previsão de tráfego pode ser feita através de modelos sazonais como:

- ARIMA Sazonal (SARIMA) (HABTEMICHAEL; CETIN, 2016);
- STARIMA Sazonal (PFEIFER; DEUTSCH, 1981; RAO; ANTUNES, 2004);
- dados em painel (BALTAGI, 2021);
- diferenciação determinística, também chamada de diferenciação sazonal, obtida pela diferenciação dos dados pela média histórica ou pelos dados do último ciclo (ALMEIDA, 2012; ZHAO et al., 2018).

Figura 2-4 – Medição de congestionamento horário (a) e diário (b) em todo Município de São Paulo



Fonte Tomtom - https://www.tomtom.com/en_gb/traffic-index/sao-paulo-traffic/

2.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO 2

Ao longo deste capítulo foram abordados os temas relativos à Estimção e à Previsão de Tráfego de curto prazo, sendo apresentados aspectos de importância aos estudos dessas matérias, dentre os quais destacaram-se: as ferramentas que podem ser utilizadas, a sazonalidade, o impacto de fatores externos, as variáveis que caracterizam o tráfego e onde é possível obtê-las, a análise espaço-temporal e a necessidade de tratamento dos dados em tempo real.

Na mesma direção, foi apresentada a aplicabilidade dos modelos de estimação em sistemas de controle, como os semaforicos, e em sistemas de simulação de tráfego.

A vasta literatura existente e a presença de várias revisões científicas relacionadas ao tema, conforme indicado no capítulo anterior, indicaram a atualidade da pesquisa e servirão de base teórica à elaboração da metodologia que será apresentada no próximo capítulo.

Foi verificada a possibilidade de uso de diversas ferramentas para uso nos modelos de estimação. Todavia, verificou-se que não existe uma linha de estudo predominante até o momento.

Também foi possível verificar a necessidade de análise e tratamento dos dados, no tocante à valores atípicos e ausência de dados, a fim de evitar o enviesamento dos resultados.

Com base na bibliografia verificada, foi possível identificar que, os modelos com Dados em Painel utilizados na área de previsão de tráfego ou se utilizam do modelo longitudinal, ou utilizam a estrutura hierárquica sem considerar especificamente a série temporal dos dados. Assim, conforme apresentaremos nos próximos capítulos, nossa pesquisa inova, ao aplicar o modelo hierárquico considerando no primeiro nível a informação da série temporal da variável dependente.

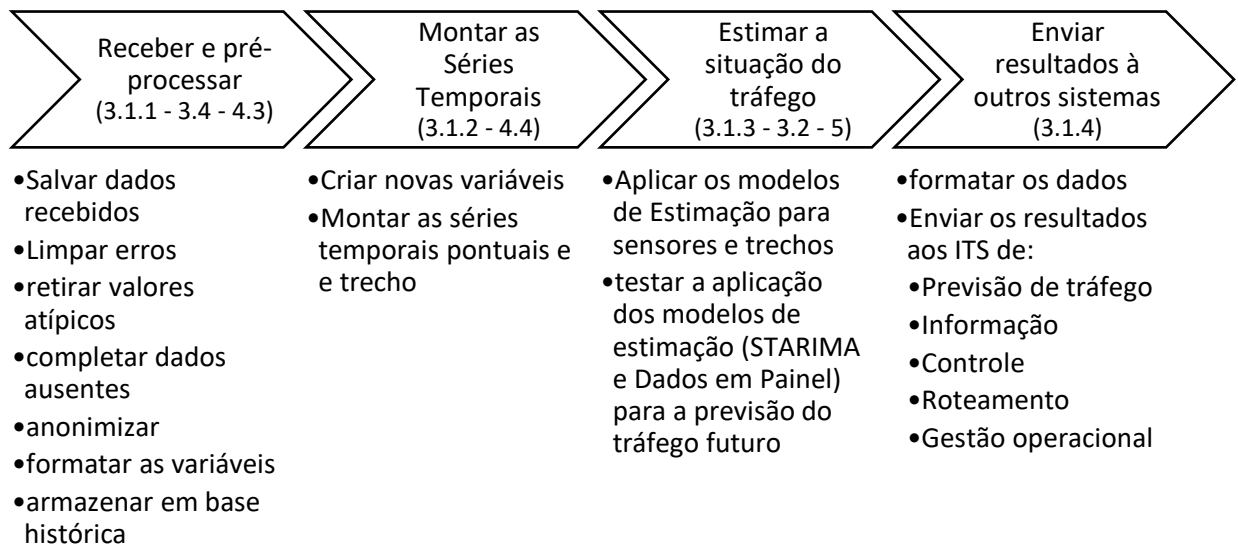
Na mesma direção inovadora, nesta tese avaliamos modelos de matrizes de pesos com o uso de modelos STARIMA, comparando diversos tipos de peso e vizinhança. O objetivo foi o de melhor entender a importância desse aspecto e de verificar quais variáveis de relacionamento trariam maior ganho à estimação.

3 METODOLOGIA

Este capítulo se inicia com a apresentação da metodologia que propomos para o processo de estimação, baseada em dados de sensores de campo e estruturada em 4 módulos. Nos itens seguintes são detalhados alguns dos componentes da metodologia. Os scripts desenvolvidos em Python e R e o aplicativo STARIMA em R/Shiny estão transcritos nos Apêndices C e D respectivamente.

O processo de estimação está baseado na metodologia descrita ao longo deste capítulo. A Figura 3-1 ilustra a metodologia e serve de base para o encadeamento dos tópicos que detalharão cada módulo. A metodologia está estabelecida na forma de um processo, em que os dados recebidos dos sensores passam por um pré-tratamento e geram as séries temporais que serão utilizadas no módulo de estimação.

Figura 3-1 – Fluxograma da metodologia proposta



Fonte: o autor

No pré-processamento e após a montagem das séries temporais é efetuada uma análise de dados, visando garantir a qualidade deles e evitar que valores atípicos prejudiquem o processo de estimação. Nesse mesmo módulo é efetuada a complementação de dados ausentes.

Somente então o método de estimação é executado e os resultados obtidos são enviados à sistemas externos, como os de previsão de tráfego, gestão do fluxo do tráfego e informação

ao usuário. Por fim, o método prevê que os dados sejam armazenados, servindo de base para a complementação de dados ausentes de intervalos seguintes e como base para a análise dos dados, recalibração periódica do modelo e em estudos futuros.

3.1 MÓDULOS COMPONENTES DA METODOLOGIA

Cada módulo da metodologia será descrito a seguir, sendo que o módulo de estimação é melhor detalhado nos itens seguintes. Salienta-se que nos módulos de pré-processamento e após a montagem das séries temporais é feita uma avaliação de dados atípicos, conforme descrito no item 4.5.

3.1.1 Módulo de recepção e pré-processamento dos dados

Cada equipamento envia aos servidores os dados registrados nos últimos 4 minutos contendo: Local, Data/Hora do registro, número do registro, Velocidade, Ocupação e Placa do Veículo. Esses dados precisam ser recebidos, armazenados e formatados de modo a permitir seu uso nos diversos passos da metodologia.

Este tratamento inicial envolve a retirada de registros corrompidos durante a transmissão, e de valores extremos. Está incluída também a anonimização dos dados, de modo a permitir sua manipulação e gravação, sem incorrer em contraposição à Lei de Sigilo de Dados.

De modo resumido temos:

- a) Dados recebidos são salvos em arquivos no formato “csv”;
- b) Valores atípicos (outliers) e dados corrompidos são retirados;
- c) Os dados são formatados no tocante à data, casas decimais, e formato da variável.

Neste passo, variáveis como nome da via, tem ou não tem faixa de ônibus, dias da semana e períodos do dia, que são variáveis categóricas, devem ser classificadas de modo adequado considerando a classe e não um eventual valor numérico atribuído à elas.

Observa-se que, para uso nos modelos com Dados em Painel, é necessário criar uma variável onde os intervalos sejam representados por um formato número que possibilite sua ordenação;

d) Dados ausentes são complementados:

Para uso nos modelos STARIMA, primeira parte da pesquisa, a complementação dos dados, foi feita de acordo com o último valor disponível. Da análise dos gráficos resultantes verificou-se que para aplicações futuras existe a necessidade de melhoria do processo, com a inclusão de estratégia para complementação dos dados ausentes quando de sua aplicação em situação de produção:

- recebimento de registro vazio: indica a ausência de passagem de veículos. São considerados Volume pontual nulo, Velocidade pontual igual à Velocidade Regulamentada, Tempo de Percurso de acordo com a velocidade regulamentada;
- informação do sensor não recebida no intervalo: no módulo seguinte os dados das séries temporais são complementados com os dados estimados no intervalo anterior. Neste caso, se os registros forem recebidos em atraso, deve ser efetuada análise quanto à presença de valores atípicos e posterior inserção na base de dados geral, podendo ser considerados nos cálculos realizados nos momentos posteriores;

e) Dados sensíveis:

- A placa é substituída por um ID: para esta tese foi necessário efetuar esta substituição logo de início, visando garantir o sigilo dos dados conforme exigência da SMMT-SP. A substituição foi feita com destruição da informação original, não sendo possível o retorno da informação da placa;
- Na aplicação prática a placa será utilizada apenas para o cálculo do tempo de percurso, sem interface homem/máquina. Assim, a substituição da placa por um ID poderá ser feita em paralelo, no momento de armazenamento dos dados, logo após sua recepção. Essa substituição poderá ser ajustada conforme análise da equipe de segurança de dados, da própria empresa detentora dos dados;

f) Dados são incluídos na base de “dados medidos” em formato csv, persistindo por um período adequado à aplicação. Nesta tese, para o desenvolvimento dos modelos foram utilizados diversos dias de dados. Todavia, na etapa de estimação, foram utilizados no máximo 8 intervalos de tempo anteriores ao de medição. A cada intervalo os dados mais antigos podem ser retirados da base de dados medidos e

são salvos em registros históricos, que servem de base para atualização periódica dos modelos.

3.1.2 Módulo de montagem das séries temporais

Com base nos dados preparados no módulo anterior devem ser elaborados bancos de dados com as séries temporais, de acordo com o modelo de estimação adotado. Na elaboração das séries temporais será adotado o passo de 4min (igual ao passo de recebimento dos dados);

Serão preparadas séries temporais locais, indicando os valores relativos aos sensores, e séries temporais de trecho, indicando os valores obtidos através do pareamento de placas entre sensores consecutivos.

O processamento dos itens pontuais e de trechos pode ser feito em paralelo.

- a) **Séries Pontuais:** para cada local de fiscalização são obtidos Volume (quantidade de registros não nulos), Volume em Veículos Equivalentes (Pesados, Automóveis, Motos), Velocidade Média (médias das velocidades registradas) e Ocupação Média (média das ocupações registradas) e Ocupação no Intervalo (soma das ocupações registradas num intervalo);
- b) **Séries de Trechos:** são calculados os Tempos de Percurso de cada trecho, através do pareamento de placas. Considera-se a mediana dos tempos de percurso dos veículos que chegam ao final do trecho, durante o período avaliado. A opção pela mediana visa eliminar erros de medição e evitar o impacto de algum veículo trafegando muito mais lento que os demais. Assim como eliminar distorções que poderiam ser causadas devido à velocidade das motocicletas, que trafegam pelo corredor entre veículos, quando o tráfego está congestionado;
- c) Como indicado, na alínea d) do módulo anterior, no caso de dados ausentes, deve ser feita a complementação através do uso dos valores estimados no intervalo anterior.

Neste módulo também devem ser criadas as variáveis explicativas que serão utilizadas nos modelos com Dados em Painel. São incluídas variáveis que caracterizam os locais, permitindo a criação de modelos hierárquicos, variáveis categóricas, variáveis multiplicativas, dados dos locais a montante e a jusante do local avaliado e séries temporais com os dados defasados. No item 4.4 apresentamos as variáveis utilizadas nos testes.

A quantidade de intervalos utilizados na defasagem das séries temporais depende do estudo realizado. Podem ser criadas variáveis defasadas de 1, 2, 3, ..., T intervalos formando uma sequência, como fizemos. Na mesma direção, podem ser criadas séries temporais defasadas de acordo com sazonalidades conhecidas como, por exemplo, o mesmo momento no dia anterior, ou o mesmo dia na semana anterior.

Nas análises espaço-temporais, com uso de Dados em Painel, são criadas variáveis com os dados das medições em locais vizinhos a montante e jusante. Todavia, sempre ocorrerão pontos ou trechos nas bordas da rede viária. A situação de borda implica que o local não tem vizinho a jusante ou a montante, podendo ocorrer a situação de um local numa via isolada, onde não existam vizinhos diretos a serem considerados. Duas alternativas são possíveis neste caso: a criação de um local vizinho fictício, fora da rede, para o qual se atribuem valores fictícios; ou a criação de modelos específicos de acordo com a existência de vizinhança. A segunda opção implica em criar 4 modelos de estimação: um modelo para os locais que tem vizinhos dos dois lados; um modelo para quem tem apenas vizinho a montante, outro para quem tem apenas vizinho a jusante e outro para locais isolados.

Ao final deste módulo, visando o desenvolvimento do modelo os dados devem ser separados em dois, formando os bancos de treino e de teste. Neste caso diversas técnicas podem ser utilizadas. Nesta tese utilizamos uma abordagem comum que é a de separar um período inicial para treino e o período final para testes.

3.1.3 Módulo de estimação

Neste módulo são aplicados os modelos de estimação, como os avaliados nesta tese.

Os modelos são testados de acordo com suas especificidades, todavia a aplicação é feita de modo a possibilitar posterior comparação dos resultados de estimação obtidos.

Os modelos são executados e comparados visando a estimação em $t+1$. De modo complementar, serão apresentadas aplicações dos modelos na previsão para horizontes futuros:

- a) Com base nas séries temporais preparadas no módulo anterior, atualizadas até $t=0$, serão efetuadas estimativas para $t+1$:
- b) Previsões para tempos posteriores como $t+5$ e $t+15$ serão efetuadas em busca de analisar a viabilidade do uso dos modelos de estimação nessa situação.

No caso dos modelos, com uso do STARIMA, especial cuidado deve ser tomado na elaboração das matrizes espaciais. Estas são fundamentais ao método, pois possibilitam a inclusão das características de relacionamento e proximidade entre os locais de estudo. O item 3.3 apresenta de modo mais detalhado o estudo desenvolvido ao longo desta tese sobre as matrizes de peso espacial. Os modelos poderiam ser selecionados com base no Critério de Informação de Akaike (AIC), todavia, por não estar disponível no aplicativo utilizado, a seleção será feita com base no MAPE e RMSE.

No caso dos Dados em Painel, diversos modelos podem ser treinados e comparados pela verossimilhança (LogLik) e pelo AIC. Note-se que, devido à quantidade de variáveis explicativas disponíveis, se torna inviável testar todos os modelos possíveis. Assim, em que pese que a inclusão ou exclusão de variáveis explicativas num modelo dependa da presença de todo o restante do modelo, é importante buscar uma estratégia de trabalho. Nesta tese utilizamos a seguinte sequência de trabalho:

- a) Iniciar pelo modelo Hierárquico Nulo para avaliar se os agrupamentos propiciam melhor previsão do que desagrupados. Neste caso não são utilizadas as variáveis explicativas;
- b) Testar a inclusão de variáveis para Efeitos Aleatórios no modelo Nulo e escolher a melhor composição e agrupamento para teste no passo “d”;
- c) Testar as variáveis no modelo Longitudinal (sem agrupamentos) apenas com o primeiro nível que considera a série histórica com os dados ordenados. Iniciaremos com a inclusão de um grande número de variáveis e a cada modelo testado serão retiradas as variáveis menos significantes;
- d) Com base nos passos anteriores serão efetuados os testes de modelos com as variáveis fixas e aleatórias. É possível que algumas variáveis, antes significativas, passem a não apresentar significância estatística. De outro lado, variáveis que foram descartadas podem voltar a ser testadas.

Qualquer que seja a ferramenta estatística utilizada, é importante testar os modelos com base no banco de teste e compará-los pelo MAPE e RMSE ou outro índice adequado ao estudo específico. Nesta tese, esta foi uma etapa importante, já que alguns modelos com bom desempenho na etapa de treino apresentaram desempenho inadequado na etapa de testes. Isso pode ocorrer no caso de certo grau de sobreajuste (overfitting).

3.1.4 Módulo de envio de dados a outros sistemas

Os dados obtidos no processo de estimação podem ser usados na retroalimentação do sistema de estimação, mas também podem ser utilizados em outros sistemas de controle e gestão do tráfego.

Os sistemas de previsão das condições do tráfego em períodos futuros, com diversos horizontes de previsão, necessitam utilizar dados atualizados para a situação atual.

Sistemas como os de controle centralizado de semáforos, roteamento de veículos, informação da qualidade do fluxo e de gestão operacional se beneficiam de informações atualizadas e de previsões das futuras condições do fluxo.

Assim, os dados finais obtidos devem ser formatados de modo a permitir seu uso em outros sistemas. Para tal, os dados devem ser organizados por sensor ou trecho e com as variáveis de acordo com os padrões mais utilizados, como velocidade em m/s ou km/h e tempos de percurso em segundos.

3.2 MODELOS DE ESTIMAÇÃO

Devido à grande variedade de modelos de estimação disponíveis, nesta tese optou-se pelos modelos de séries temporais STARIMA e de Regressões com Dados em Painel. Estes foram escolhidos por permitem a elaboração de modelos espaço-temporais e por não serem do tipo “caixa preta”, permitindo melhor entendimento dos fenômenos estudados. Também pesou, na escolha dos modelos, o fato de ambos possibilitarem a implantação escalonada, permitindo a aplicação por região ou corredor de tráfego. A escalabilidade do modelo torna possível sua implementação em cidades de grande porte, sendo que a quantidade de locais utilizada em cada grupo dependerá das características operacionais, da ferramenta utilizada e da capacidade computacional utilizada.

Outro aspecto importante é que estes modelos se apresentam com menor quantidade na literatura avaliada, de modo que, por serem menos explorados, existe a possibilidade desta tese contribuir significativamente no aumento da base de conhecimento, relativa à aplicação deles na elaboração de modelos de estimação de tráfego.

3.2.1 Modelo autorregressivo - STARIMA

Trata-se de um modelo puramente indutivo para o estudo espaço temporal de séries temporais, que pode ser aplicado em ST-STTF, quando as condições de tráfego são ou podem ser transformadas em estacionárias. O modelo STARIMA permite prever as condições da rede completa com apenas um modelo, independentemente do número de locais de medição do tráfego (KAMARIANAKIS; PRASTACOS, 2005), com precisão relativamente alta e número de parâmetros do modelo relativamente baixo (MIN et al., 2009).

Para aplicação do modelo STARIMA é assumida a estacionariedade e constância dos efeitos em relação ao tempo, isto é, o efeito de um incidente em um local é repassado aos demais locais de sua vizinhança, a partir do instante em que ocorreu. Assim, apresenta limitações na falta de ligação com a realidade de algumas situações devido à essas hipóteses. O modelo expressa cada observação em um instante t e local r como uma combinação linear ponderada de observações passadas. O mecanismo básico para essa representação é o ordenamento hierárquico dos vizinhos de cada local que resulta numa sequência correspondente de matrizes ponderadas (KAMARIANAKIS; PRASTACOS, 2005).

O modelo STARIMA (6) pode ser entendido como o resultado da conjunção de uma série de modelos de previsão, que inclui análise Espaço Temporal (ST), Autorregressão (AR), Diferenciação/Integração (I) e Média Móvel (MA) – Expressão 3-1:

$$y_i^{(d)}(t) = \underbrace{\sum_{k=1}^p \sum_{l=0}^{\lambda_k} \beta_{kl} \cdot W^{(l)} \left(y_i^{(d)}(t-k) \right)}_{\text{Autorregressão}} - \underbrace{\sum_{k=1}^q \sum_{l=0}^{m_k} \alpha_{kl} \cdot W^{(l)} (\varepsilon_i(t-k))}_{\text{Média Móvel}} + \varepsilon_i(t) \quad (3-1)$$

Diferenciação
Diferenciação

Onde $y_i^{(d)}(t)$ refere-se ao incremento de fluxo de tráfego no local, num determinado instante t , sendo d o número de diferenciações; λ_k denota a ordem espacial no l -ésimo termo autorregressivo da ordem temporal da autorregressão, variando de 1 a p ; m_k denota a ordem espacial do k -ésimo termo da média móvel (que varia de 1 a q , o termo temporal da média móvel); $W^{(l)}$ é a matriz $N \times N$ de pesos espaciais de ordem l ; ε_{t-k} e ε_t são os erros. Os parâmetros autorregressivos β_{kl} e de média móvel α_{kl} são determinados de modo iterativo. Sendo N a quantidade de pontos de medição, k é o *lag* temporal, l o *lag* espacial.

No desenvolvimento de modelos baseados no STARIMA, é necessário montar séries temporais da variável a ser avaliada. Para possibilitar o uso do aspecto espacial, é necessário que os dados sejam organizados de modo que seja criada uma tabela com a série temporal de cada local colocada em uma coluna Tabela 3-1.

Tabela 3-1 – Organização dos dados para aplicação em modelos baseados em STARIMA

Intervalo	Local 1	Local 2	...	Local /
t=1	$y_1(1)$	$y_2(1)$...	$y_l(1)$
t=2	$y_1(2)$	$y_2(2)$...	$y_l(2)$
...
t=T	$y_1(T)$	$y_2(T)$...	$y_l(T)$

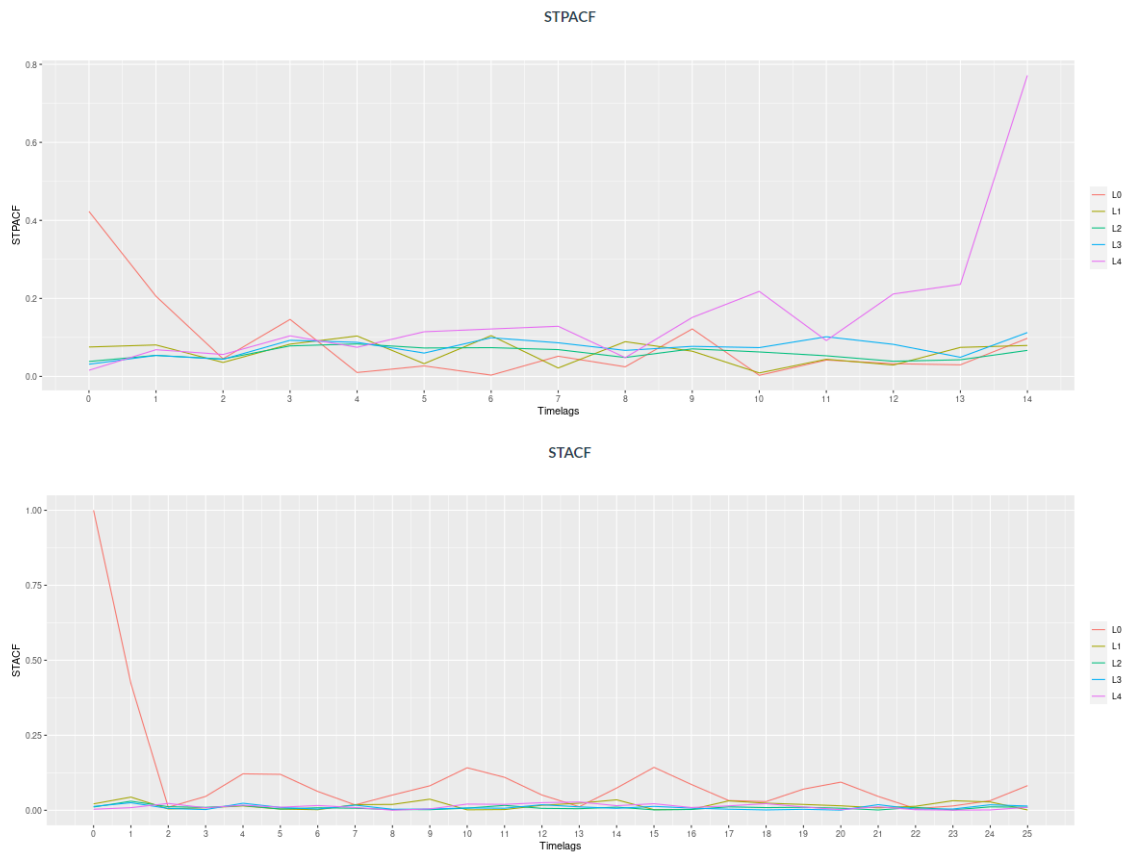
Fonte: do autor

Para sua aplicação em modelos de previsão, são necessárias as fases de identificação do modelo (definição dos 5 parâmetros essenciais p , m , d , q e λ), estimação dos parâmetros autorregressivo e de média móvel β_{kl} e α_{kl} e calibração. Para maiores detalhes de uso do modelo sugerimos Pfeifer e Deutsch (1980) e Rao e Antunes (2004)

Nesta tese, a construção do modelo é feita com base em Pfeifer e Deutsch (1980a), sendo que o aplicativo, descrito no Apêndice D, define o valor da ordem de integração d necessária para garantir a estacionariedade das séries temporais. Os valores diferenciados são armazenados, de modo que ao final dos cálculos eles são “integrados” ao resultado, possibilitando a obtenção dos dados estimados.

A escolha inicial das ordens da autocorrelação q e m é feita a partir de inspeção visual da aplicação da função STACF (Space Time Autocorrelation Function) e as ordens de p e λ , da média móvel, baseiam-se na STPACF (Space Time Partial Autocorrelation Function). Do exemplo da Figura 3-2 resultariam os valores de: $q=1$ verificado no decaimento linha vermelha (L0) na ordem 2; $m=4$, devido presença de variação na linha lilás (L4) que indica a ordem 4; $p=1$ pelo decaimento da autocorrelação indicada pela linha vermelha e $\lambda=0$ por ausência de variação que se destaque nas demais ordens.

Figura 3-2 – Gráficos STPACF e ATACF



Fonte: elaborado pelo autor com uso do aplicativo STARIMA (PAZ; TRINTINALIA, 2021)

Observa-se que, devido ao método de determinação das ordens espaço temporais ser feito de modo visual, existe a possibilidade de erro por parte do pesquisador na escolha. Nesse sentido, é possível efetuar o processo seguinte com base em mais de uma composição de valores para as ordens e escolher a composição que apresentar melhor acurácia na etapa de treino ou testes.

A definição final das ordens será feita através de um procedimento de análise exaustiva de todos os modelos possíveis, tendo como limite máximo as ordens de $p+1$, $\lambda+1$, $q+1$ e $m+1$, sendo p , λ , q e m as ordens obtidas nos gráficos. Por fim, os parâmetros β_{kl} e α_{kl} são dados pelo aplicativo a partir da aplicação do método de Levenberg Marquadt. No entanto, é importante perceber que os valores das ordens do modelo determinam o número de parâmetros a ser estimado no passo seguinte, que pode ser obtido pelas Equações 3-2 e 3-3:

$$N_{\alpha} = (m + 1) \cdot q \quad (3-2)$$

$$N_{\beta} = (\lambda + 1) \cdot p \quad (3-3)$$

Onde N_α representa a quantidade de parâmetros da média móvel a serem obtidos e N_θ a quantidade de parâmetros da autocorrelação

3.2.2 Modelos de regressão de Dados em Painel

Os Modelos de Regressão de Dados em Painel podem ser classificados como Longitudinais e Multinível (Hierárquicos).

Nos modelos Longitudinais os indivíduos são medidos em diversos instantes, criando-se diversas *cross-sections* que são organizadas em forma de séries temporais, sendo uma para cada intervalo de medição. Os modelos Longitudinais podem ser definidos pela Expressão 3-4 (FÁVERO; BELFIORE, 2017):

$$Y_{at} = c_a + \sum_{g=1}^G b_g * X_{gat} + e_{at} \quad (3-4)$$

Sendo que Y_{at} é a variável dependente para o indivíduo a no intervalo t ; c é o intercepto para o indivíduo a ; b é o coeficiente das g -ésimas variáveis explicativas; X é a g -ésima variável explicativa. O erro idiosincrático é denominado como e .

A Tabela 3-2 apresenta um exemplo de organização dos dados das *cross-sections* visando sua aplicação em modelos Longitudinais com Dados em Painel.

Tabela 3-2 – Organização dos dados para aplicação em modelos Longitudinais com Dados em Painel

Observação	Indivíduo (a)	Intervalo (t)	Y_{at}	X_{1at}	X_{2at}	X_{Gat}
1	1	1	Y_{1t1}	$X_{1\ 1\ t1}$	$X_{2\ 1\ t1}$	$X_{G\ 1\ t1}$
2	1	2	Y_{1t2}	$X_{1\ 1\ t2}$	$X_{2\ 1\ t2}$	$X_{G\ 1\ t2}$
...
...	1	T	Y_{1tT}	$X_{1\ 1\ tT}$	$X_{2\ 1\ tT}$	$X_{G\ 1\ tT}$
...	2	1	Y_{2t1}	$X_{1\ 2\ t1}$	$X_{2\ 2\ t1}$	$X_{G\ 2\ t1}$
...	2	2	Y_{2t2}	$X_{1\ 2\ t2}$	$X_{2\ 2\ t2}$	$X_{G\ 2\ t2}$
...
...	2	T	Y_{2tT}	$X_{1\ 2\ tT}$	$X_{2\ 2\ tT}$	$X_{G\ 2\ tT}$
...
N	A	T	$Y_{A\ tT}$	$X_{1\ A\ tT}$	$X_{2\ A\ tT}$	$X_{G\ A\ tT}$

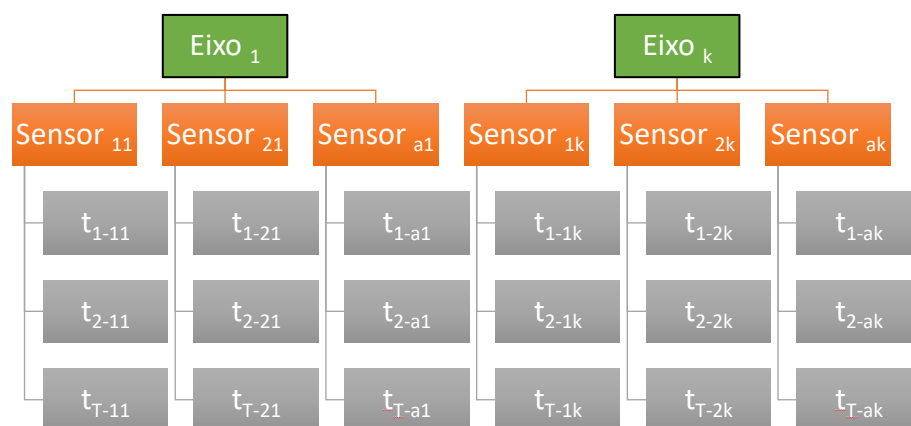
Fonte: adaptado de (FÁVERO; BELFIORE, 2017)

Segundo Fávero e Belfiore (2017), os Modelos Hierárquicos Lineares (HLM) e os Modelos Hierárquicos Não Lineares (HNM) ou, genericamente, Modelos Multinível de Regressão para Dados em Painel, possibilitam a análise das relações entre os diversos níveis, o que permite um maior entendimento do problema de estudo.

Diferentemente do que ocorre nos modelos de regressão não hierárquicos, o modelo hierárquico nos possibilita avaliar o impacto de cada variável explicativa em cada um dos agrupamentos e explorar assim seu efeito explicativo com relação à variável dependente.

O modelo hierárquico deve ser estabelecido através do agrupamento de indivíduos e do agrupamento de agrupamentos, nos níveis superiores. Assim, a estrutura dos dados deve possibilitar a organização em forma de árvores, não sendo necessariamente balanceadas, de modo que é possível que ocorram agrupamentos com mais indivíduos do que outros. A Figura 3-3 exemplifica uma das estruturas que será utilizada nesta tese, sendo o nível 1 composto pelas medidas nos intervalos e representada por t_{T-ak} , os a_k sensores estão no segundo nível e o terceiro nível é composto pelos k eixos viários estudados.

Figura 3-3 – Exemplo de estrutura hierárquica por local



Fonte: do autor

Dessa forma, o modelo obtido apresenta valores específicos, para cada agrupamento efetuado em cada um dos níveis, possibilitando maior acurácia na previsão de dados futuros.

Para a aplicação dos modelos baseados em Dados em Painel, O padrão de entrada de dados para os modelos em painel deve ser aquele em que as variáveis são colunas e as medições para cada intervalo t de cada indivíduo a de cada grupo k estão descritas nas linhas, conforme exemplificado na Tabela 3-3.

Tabela 3-3 – Organização dos dados para aplicação em modelos Hierárquicos com Dados em Painel

Intervalo (t) (Nível 1)	Indivíduo (a) (Nível 2)	Grupo (k) (Nível 3)	Y_{tak}	X_{1ak}	X_{2ak}	X_{Gak}	W_{1k}	W_{2k}	W_{Sk}
1	1	1	$Y_{t1\ 11}$	X_{111}	X_{211}	X_{G11}	W_{11}	W_{21}	W_{S1}
2	1	1	$Y_{t2\ 11}$	X_{111}	X_{211}	X_{G11}	W_{11}	W_{21}	W_{S1}
...							
T	1	1	$Y_{tT\ 11}$	X_{111}	X_{211}	X_{G11}	W_{11}	W_{21}	W_{S1}
1	2	1	$Y_{t1\ 21}$	X_{121}	X_{221}	X_{G21}	W_{11}	W_{21}	W_{S1}
2	2	1	$Y_{t2\ 21}$	X_{121}	X_{221}	X_{G21}	W_{11}	W_{21}	W_{S1}
...							
T	2	1	$Y_{tT\ 21}$	X_{121}	X_{221}	X_{G21}	W_{11}	W_{21}	W_{S1}
1	3	2	$Y_{t1\ 32}$	X_{132}	X_{232}	X_{G22}	W_{12}	W_{22}	W_{S2}
2	3	2	$Y_{t2\ 32}$	X_{132}	X_{232}	X_{G22}	W_{12}	W_{22}	W_{S2}
...							
T	3	2	$Y_{tT\ 32}$	X_{132}	X_{232}	X_{G22}	W_{12}	W_{22}	W_{S2}
...							
T	A	K	$Y_{tT\ AK}$	X_{1AK}	X_{2AK}	X_{GAK}	W_{1K}	W_{2K}	W_{SK}

Fonte: adaptado de (FÁVERO; BELFIORE, 2017)

Nos estudos com modelos hierárquicos lineares utiliza-se como primeira abordagem o HLM Nulo que se caracteriza pela ausência de variáveis explicativas, tanto nos Efeitos Fixos como nos Efeitos Aleatórios, e pela organização dos dados em diversos níveis de agrupamento, também chamados de níveis de aninhamento (FÁVERO; BELFIORE, 2017)– Expressões 3-5:

$$\text{Nível 1 } Y_{tak} = \pi_{0ak} + e_{tak}$$

$$\text{Nível 2 } \pi_{0ak} = b_{00k} + r_{0ak}$$

$$\text{Nível 3 } b_{00k} = \gamma_{000} + u_{00k}$$

$$\text{HLM3 Nulo com medidas repetidas: } Y_{tak} = \gamma_{000} + u_{00k} + r_{0ak} + e_{tak} \quad (3-5)$$

Onde Y_{tak} é a variável dependente para o indivíduo a do grupo k no intervalo t , π , b e γ são os coeficientes dos níveis 1, 2 e 3 respectivamente. Os erros idiossincráticos de primeiro, segundo e terceiro nível são respectivamente, e , r e u .

Com o uso do modelo nulo é possível verificar quais dos níveis existentes são estatisticamente significantes e testar inicialmente o impacto da autocorrelação e do impacto das médias móveis nos resíduos. Assim, a análise dos modelos nulos permite otimizar o

processo seguinte, que consiste na inclusão de variáveis explicativas tanto no campo dos Efeitos Fixos como no de Efeitos Aleatórios.

Quando da utilização de HLM com medidas repetidas ao longo do tempo, como nesta tese, é necessário que os intervalos de medição sejam incluídos no primeiro nível do modelo, onde a variável explicativa “*intervalo*” representa a medida repetida. Os níveis 2 e 3 representam os outros dois níveis de agrupamento hierárquico (FÁVERO; BELFIORE, 2017) – Expressões 3-6:

$$\begin{aligned}
 \text{Nível 1} \quad Y_{tak} &= \pi_{0ak} + \pi_{fak} * \text{intervalo}_{ak} + e_{tak} \\
 \text{Nível 2} \quad \pi_{fak} &= b_{f0k} + \sum_{g=1}^{G_f} b_{fgk} * X_{gak} + r_{fak} \\
 \text{Nível 3} \quad b_{fgk} &= \gamma_{fg0} + \sum_{s=1}^{S_{fg}} \gamma_{fgs} * M_{sk} + u_{fgk} \quad (3-6)
 \end{aligned}$$

Onde Y_{tak} é a variável dependente para o indivíduo a do grupo k no intervalo t , π , b e γ são os coeficientes dos níveis 1, 2 e 3 respectivamente; X é a g -ésima variável explicativa do nível 2 e M é a s -ésima variável explicativa do nível 3. Salienta-se que a variável X é invariante para os indivíduos a , no instante t , e M é invariante em t para a no grupo k . Por fim, e , r e u são os erros idiossincráticos de primeiro, segundo e terceiro nível.

Para os modelos não lineares (HNM), aplicados à dados de contagem, são utilizadas as Expressões 3-7 (FÁVERO; BELFIORE, 2017) onde a expressão de nível 1 é alterada para considerar a distribuição e Poisson. Neste caso os cálculos envolvem a análise de dados em formato exponencial, portanto não linear. Assim, sendo φ_{tak} o número esperado, é necessário efetuar a transformação logarítmica para obter o valor da variável dependente.

$$\begin{aligned}
 \text{Nível 1} \quad Y_{tak} &= \ln(\varphi_{tak}) = \pi_{fak} + \pi_{1ak} * \text{intervalo}_{ak} \\
 \text{Nível 2} \quad \pi_{fak} &= b_{f0k} + \sum_{g=1}^{G_f} b_{fgk} * X_{gak} + r_{fak} \\
 \text{Nível 3} \quad b_{fgk} &= \gamma_{fg0} + \sum_{s=1}^{S_{fg}} \gamma_{fgs} * M_{sk} + u_{fgk} \quad (3-7)
 \end{aligned}$$

3.3 MATRIZES DE PESO ESPACIAL

Como detalhado no capítulo anterior as matrizes espaciais são utilizadas em diversos modelos de análise espaço temporal. Nesta tese as matrizes são implementadas nos modelos

de estimação STARIMA. Para isso é efetuada uma análise detalhada dos pesos e vizinhanças que podem ser empregadas. A aplicação de matrizes de peso nos modelos baseados em dados em painel também é possível, no entanto, não será efetuada nesta tese pois, neste caso, optou-se pela construção de um banco de dados já estruturado com as informações dos locais a montante e a jusante.

As matrizes espaciais do modelo STARIMA foram construídas de acordo com a “ordem” de proximidade entre os vizinhos (vizinhança/proximidade) e a intensidade de relacionamento (distância/pesos) com base nas variações apresentadas na Tabela 3-4.

Todos os cenários consideram o impacto do fluxo à montante e à jusante. Os 3 tipos assimétricos propostos buscam incorporar a diferença do estado do fluxo do trecho, quando está posicionado à montante e à jusante do ponto avaliado.

Tabela 3-4 - Características das matrizes utilizadas nos testes

Tipo	Descrição	Peso à Montante	Peso à Jusante	Vizinhança			Simétrica	
				VC	LP	AG	Sim	Não
UN	Contiguidade Unitária	1	1	E	E		x	
1/D	Inverso da Distância	$\frac{1}{D_{nm\ nj}^j}$	$\frac{1}{D_{nm\ nj}^j}$	E	E	ER	x	
1/T	Inverso do Tempo de Percurso	$\frac{1}{T_{nm\ nj}^j}$	$\frac{1}{T_{nm\ nj}^j}$	E	E	ER	x	
1-1/T	Inverso do Tempo de Percurso e seu Complemento	$\frac{1}{T_{nm\ nj}^j}$	$1 - \frac{1}{T_{nm\ nj}^j}$	ER	ER	ER		x
1/V	Inverso da Velocidade	$\frac{1}{V_{nm\ nj}^j}$	$\frac{1}{V_{nm\ nj}^j}$	E	E	ER	x	
1-1/V	Inverso da Velocidade e seu complemento	$1 - \frac{1}{V_{nm\ nj}^j}$	$\frac{1}{V_{nm\ nj}^j}$	ER	ER	ER		x
1/L	Inverso do número de LAGs de percurso	$\frac{1}{L_{nm\ nj}^j}$	$\frac{1}{L_{nm\ nj}^j}$	E	E	ER	x	
1-1/L	Inverso do número de LAGs de percurso e seu Complemento	$\frac{1}{L_{nm\ nj}^j}$	$1 - \frac{1}{L_{nm\ nj}^j}$	E	E	ER		x

(E= teste apenas nos 2 Eixos; ER = testes nos eixos e na rede completa)

Fonte: O autor

Na Tabela 3-4, cada tipo de matriz é descrita com relação a como são calculados os pesos utilizados, de modo que a coluna “Peso à Montante” indica o valor de $w_{nm\ nj}^l$ a ser utilizado na linha nj e coluna nm , já o “Peso à Jusante” deve ser utilizado na linha nm e coluna nj . No item 3.3.2 será apresentado um exemplo teórico dessa utilização. A característica de uma matriz assimétrica é a utilização de pesos diferentes nas colunas de peso à montante e jusante.

Para os três tipos de Vizinhança – por Contiguidade (VC), por Tempo de Percurso (LP) e Agrupado (AG) - apresentados no item 2.3.2, estão indicados os ambientes para avaliação, sendo indicado *E* para os tipos que serão avaliados apenas para os eixos de tráfego e *ER* para os que serão avaliados tanto nos eixos como com a rede de 18 sensores. Observa-se que não foi atribuído o tipo AG para as matrizes unitárias (UN), pois não faria sentido uma matriz inteiramente preenchida com valores igual a 1, já que resultaria na exclusão da característica espacial do modelo.

Um exemplo da montagem para tipo *1-1/V* é apresentado de modo detalhado no item 3.3.2. Antes disso, apresentamos no próximo item uma descrição dos tipos de peso utilizados.

3.3.1 Tipos de pesos utilizados nos modelos

A seguir descrevemos com maior detalhe cada um dos 8 tipos de peso utilizados nesta tese.

A matriz UN utiliza o valor unitário para os locais vizinhos, de acordo com a ordem de vizinhança. Assim, todos os vizinhos de mesma ordem de vizinhança possuem o mesmo peso em relação ao local avaliado. As matrizes de peso unitário são iguais às matrizes geradoras apresentadas no próximo item.

Na matriz 1/D a “Lei de Tobler” é diretamente aplicada através do uso do inverso da distância entre os pontos, de modo que quanto maior a distância entre os pontos menor é sua interrelação.

A matriz 1/T considera o inverso do tempo de percurso no link, medido em segundos, de modo que, quanto menor o tempo de percurso pressupõe-se maior correlação entre os locais de início e término do trecho, do que decorre que um maior peso deve ser atribuído.

A matriz 1/L é uma abordagem mais abrangente e simplificada para considerar o Tempo de Percurso, sendo utilizada a quantidade de intervalos (*lags*) que correspondem ao tempo de percurso em determinado trecho. Assim, considerando intervalos de 4 minutos, trechos percorridos em menos de 4 minutos recebem o valor 1, entre 4 e 8 minutos o valor 1/2, etc. Neste caso, pequenas variações na velocidade são assimiladas sem necessidade de alteração dos pesos utilizados.

A matriz 1/V é similar à do tempo de percurso, o inverso da velocidade média de percurso no trecho passa a ser considerada como variável de medição da proximidade entre os pontos.

Quanto maior a velocidade média de percurso pressupõe-se menor correlação entre os pontos de início e término do trecho.

Os modelos assimétricos testados utilizaram o artifício de apropriar o peso correspondente ao valor complementar dos modelos simétricos, em parte das células da matriz.

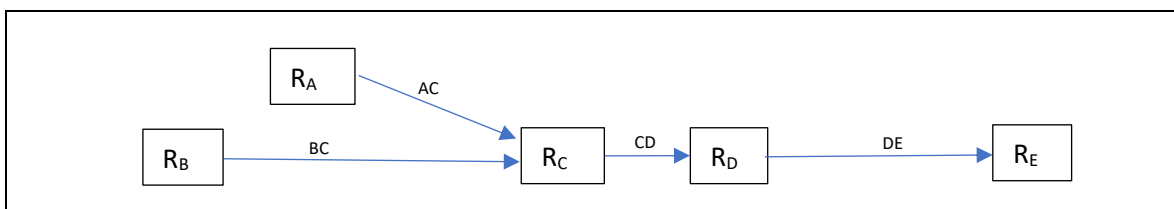
Nas matrizes 1-1/T e 1-1/L, considera-se que um tempo de percurso maior no trecho à jusante também pode estar relacionado à um maior impacto no ponto avaliado. Tempos maiores podem indicar piora no fluxo do trecho à frente, o que geralmente indica redução na capacidade de absorver o fluxo que chega do trecho anterior. Assim, nas células dos trechos à jusante é considerado o valor complementar do inverso do tempo $(1 - \frac{1}{T_{nm}^j})$, de modo que quanto maior o tempo de percurso, maior será o peso atribuído. O oposto ocorre no trecho à montante, onde tempos maiores podem indicar maior dificuldade na chegada dos veículos de modo que o peso $(\frac{1}{T_{nm}^j})$ permanece sendo utilizado.

De modo semelhante, na matriz 1-1/V foi empregado o complementar nos trechos de montante $(1 - \frac{1}{V_{nm}^j})$ o que indica maior peso no caso de maiores velocidades nesse trecho, já que velocidades maiores no trecho à montante podem indicar maior fluidez resultando em maior facilidade dos veículos em atingirem o ponto analisado. Para os trechos de jusante o peso é o mesmo do tipo 1/V, ou seja $\frac{1}{V_{nm}^j}$

3.3.2 Organização das matrizes de peso de acordo com as vizinhanças

Para melhor ilustrar a metodologia utilizaremos um exemplo fictício baseado na rede viária apresentada na Figura 3-4 cujos links são unidirecionais e estão definidos pela posição dos radares R_r , sendo possível verificar que R_A e R_B estão à montante de R_C , que está à montante de R_D , que por sua vez está à montante de R_E .

Figura 3-4 – Rede viária teórica



Fonte: o autor

Inicialmente foram elaboradas as Matrizes de Peso Gerais (Figura 3-5) para cada um dos tipos indicados na Tabela 3-4.

Em seguida, foram elaboradas as matrizes finais, efetuando-se o Produto de Hadamard (no qual cada célula de uma matriz é multiplicada pela mesma célula da outra matriz) da Matriz de Peso Geral pelas matrizes geradoras de cada tipo, AG, VC e LP.

Figura 3-5 - Matriz de Peso Geral da Rede Teórica com pesos do tipo 1-1/V

$$\begin{pmatrix} 1 & \frac{1}{V_{AB}} & \frac{1}{V_{AC}} & \frac{1}{V_{AD}} & \frac{1}{V_{AE}} \\ 1 - \frac{1}{V_{AB}} & 1 & \frac{1}{V_{BC}} & \frac{1}{V_{BD}} & \frac{1}{V_{BE}} \\ 1 - \frac{1}{V_{AC}} & 1 - \frac{1}{V_{BC}} & 1 & \frac{1}{V_{CD}} & \frac{1}{V_{CE}} \\ 1 - \frac{1}{V_{AD}} & 1 - \frac{1}{V_{BD}} & 1 - \frac{1}{V_{CD}} & 1 & \frac{1}{V_{DE}} \\ 1 - \frac{1}{V_{AE}} & 1 - \frac{1}{V_{BE}} & 1 - \frac{1}{V_{CE}} & 1 - \frac{1}{V_{DE}} & 1 \end{pmatrix}$$

Fonte: o autor

Com base na teoria dos grafos é possível determinar a Matriz de Ordens de Vizinhança por Contiguidade da rede viária fictícia apresentada na Figura 3-6. Perceba-se que R_A e R_B não estão diretamente relacionados, de modo que na matriz de vizinhança deve-se apontar valor nulo para esse par de locais. Às vizinhanças a montante atribuímos o índice m e nas de jusante o índice j .

Figura 3-6 - Matriz de Ordens de Vizinhança por Contiguidade

	R_A	R_B	R_C	R_D	R_E
R_A	0	0	1j	2j	3j
R_B	0	0	1j	2j	3j
R_C	1m	1m	0	1j	2j
R_D	2m	2m	1m	0	1j
R_E	3m	3m	2m	1m	0

Fonte: o autor

A Matriz de Ordens de Vizinhança por contiguidade, apresentada na Figura 3-7, também pode ser expressa na forma de Matrizes Geradoras, de acordo com a ordem de vizinhança por contiguidade.

Importante salientar que, em todos os tipos testados, é feita a aplicação da matriz de ordem 0 (zero) W^0 sendo esta uma Matriz Identidade. Assim, com W^0 o local avaliado influencia a si mesmo e com W^1 os locais vizinhos influenciam o local avaliado

Figura 3-7 - Matrizes geradoras das matrizes de peso por ordem para vizinhança por contiguidade

Ordem 0	Ordem 1	Ordem 2	Ordem 3
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$

Obs.: A matriz de ordem 0 é uma matriz identidade e está presente em todos os modelos testados

Fonte: o autor

Efetuando-se o Produto de Hadamard das matrizes geradoras pela Matriz de Pesos Geral é possível obter as matrizes de peso das diversas ordens para os diversos tipos de peso propostos. A Figura 3-8 apresenta como seriam as matrizes de Ordem 1, 2 e 3 do tipo 1-1/V para o exemplo apresentado, lembrando que a matriz de ordem 0 é uma matriz identidade.

Figura 3-8 – Exemplo de Matriz de peso da Rede Teórica com vizinhança por contiguidade

Ordem 1	Ordem 2	Ordem 3
$\begin{pmatrix} 0 & 0 & \frac{1}{V_{AC}} & 0 & 0 \\ 0 & 0 & \frac{1}{V_{BC}} & 0 & 0 \\ 1 - \frac{1}{V_{AC}} & 1 - \frac{1}{V_{BC}} & 0 & \frac{1}{V_{CD}} & 0 \\ 0 & 0 & 1 - \frac{1}{V_{CD}} & 0 & \frac{1}{V_{DE}} \\ 0 & 0 & 0 & 1 - \frac{1}{V_{DE}} & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & \frac{1}{V_{AD}} & 0 \\ 0 & 0 & 0 & \frac{1}{V_{BD}} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{V_{CE}} \\ 1 - \frac{1}{V_{AD}} & 1 - \frac{1}{V_{BD}} & 0 & 0 & 0 \\ 0 & 0 & 1 - \frac{1}{V_{CE}} & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{V_{AE}} \\ 0 & 0 & 0 & 0 & \frac{1}{V_{BE}} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 - \frac{1}{V_{AE}} & 1 - \frac{1}{V_{BE}} & 0 & 0 & 0 \end{pmatrix}$

Fonte: o autor

As matrizes LP consideram simultaneamente a relação espacial entre os diversos locais e o tempo de percurso entre esses locais para então estabelecer a “ordem” da matriz a ser utilizada. Assim, nas matrizes de LP de ordem espacial L_{LP} incluem-se todos os locais nos quais é possível um veículo passar durante o intervalo Δt adotado. A ordem é definida pelo $L_{nm\ nj}^j$, de acordo com o tempo de percurso – Expressão 3-8:

$$L_{nm\ nj}^j = \frac{T_{nm\ nj}^j}{\Delta t} + 1, \text{ sendo } L_{nm\ nj}^j = \text{inteiro} \quad (3-8)$$

Seguindo com o exemplo, considerando $\Delta t = 4min$ e os tempos de trajeto hipotéticos, apresentados na Tabela 3-5, e utilizando-se a teoria dos grafos, podemos remontar a Matriz de Ordens de Vizinhaça baseada no modelo LP, Figura 3-9.

Tabela 3-5 – Tempo de percurso e Ordem de Vizinhaça LP nos trechos do exemplo teórico

Trecho	AC	BC	CD	DE	ACD	ACDE	BCD	BCDE	CDE
Tempo de Percurso (min)	3,5	5,2	2,3	7,1	5,8	12,9	7,8	14,9	9,4
L_{nm}^j	1	2	1	2	2	4	3	4	3

Fonte: o autor

Figura 3-9 - Matriz de Ordens de Vizinhaça baseada no modelo LP

	R_A	R_B	R_C	R_D	R_E
R_A	0	0	1j	2j	4j
R_B	0	0	2j	3j	4j
R_C	1m	2m	0	1j	3j
R_D	2m	3m	1m	0	2j
R_E	4m	4m	3m	2m	0

Fonte: o autor

Novamente, as Matrizes Geradoras de ordens do modelo LP são obtidas com base na teoria dos grafos e nas ordens indicadas na Figura 3-9. Efetuando-se o Produto de Hadamard das matrizes geradoras LP pela Matriz de Peso Geral, obtemos as matrizes de peso das diversas ordens para os diversos tipos de peso propostos na Tabela 3-4. A Figura 3-10 apresenta como seriam as matrizes de Ordem 1, 2, 3 e 4 do tipo 1-1/V do exemplo.

Figura 3-10 - Exemplo de Matriz de peso da Rede Teórica com vizinhança LP

Ordem 1	Ordem 2
$\begin{pmatrix} 0 & 0 & \frac{1}{V_{AC}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 - \frac{1}{V_{AC}} & 0 & 0 & \frac{1}{V_{CD}} & 0 \\ 0 & 0 & 1 - \frac{1}{V_{CD}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & \frac{1}{V_{AD}} & 0 \\ 0 & 0 & \frac{1}{V_{BC}} & 0 & 0 \\ 0 & 1 - \frac{1}{V_{BC}} & 0 & 0 & 0 \\ 1 - \frac{1}{V_{AD}} & 0 & 0 & 0 & \frac{1}{V_{CE}} \\ 0 & 0 & 0 & 1 - \frac{1}{V_{CE}} & 0 \end{pmatrix}$
Ordem 3	Ordem 4
$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{V_{BD}} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{V_{CE}} \\ 0 & 1 - \frac{1}{V_{BD}} & 0 & 0 & 0 \\ 0 & 0 & 1 - \frac{1}{V_{CE}} & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{V_{AE}} \\ 0 & 0 & 0 & 0 & \frac{1}{V_{BE}} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 - \frac{1}{V_{AE}} & 1 - \frac{1}{V_{BE}} & 0 & 0 & 0 \end{pmatrix}$

Fonte: o autor

Nas matrizes de peso Agrupadas foram reunidos os pesos das correlações até a 4ª ordem de vizinhança, por contiguidade. Assim, seguindo com o exemplo teórico, a matriz de pesos agrupados para o tipo 1-1/V pode ser descrita como apresentado na Figura 3-11:

Figura 3-11 - Matriz de Pesos Agrupada da Rede Teórica com pesos do tipo 1-1/V

$$\begin{pmatrix} 0 & 0 & \frac{1}{V_{AC}} & \frac{1}{V_{AD}} & \frac{1}{V_{AE}} \\ 0 & 0 & \frac{1}{V_{BC}} & \frac{1}{V_{BD}} & \frac{1}{V_{BE}} \\ 1 - \frac{1}{V_{AC}} & 1 - \frac{1}{V_{BC}} & 0 & \frac{1}{V_{CD}} & \frac{1}{V_{CD}} \\ 1 - \frac{1}{V_{AD}} & 1 - \frac{1}{V_{BD}} & 1 - \frac{1}{V_{CD}} & 0 & \frac{1}{V_{DE}} \\ 1 - \frac{1}{V_{AE}} & 1 - \frac{1}{V_{BE}} & 1 - \frac{1}{V_{CE}} & 1 - \frac{1}{V_{DE}} & 0 \end{pmatrix}$$

Fonte: o autor

Diferentes estudos utilizam matrizes de peso espacial normalizadas, onde as linhas são tratadas para resultarem numa soma de valor um. Nesta tese essa normalização alteraria as características de proximidade adotadas na determinação dos pesos, fazendo com que, por exemplo, numa matriz simétrica um mesmo trecho tivesse pesos diferentes quando colocado em linhas diferentes da matriz. Isso decorre das composições de vizinhança adotadas nas redes de tráfego que, no caso desta tese, consistem em trajetos unidirecionais. Dessa forma, as matrizes aqui utilizadas não foram normalizadas.

3.4 ANÁLISE DOS DADOS DO TRÁFEGO

Qualquer sistema de monitoramento é sujeito a falhas tanto na fase de obtenção das informações como no seu registro e transmissão dos dados obtidos. Destas falhas podem decorrer valores atípicos (*outliers*) e ausência de dados.

Em aplicações que utilizam dados de sensores ou de veículos interconectados à infraestrutura de ITS (VxI), os dados não podem ser carregados diretamente nas camadas de otimização e controle. Existe a necessidade de efetuar um pré-tratamento dos dados incluindo a filtragem, agregação e processamento dos dados recebidos. Este processo ocorre com o objetivo de determinar e estimar os valores que devem ser aplicados para um determinado horizonte de otimização (RONCOLI; PAPAMICHAIL; PAPAGEORGIOU, 2016).

No ambiente de tráfego a ocorrência de incidentes (acidentes, obras, veículos quebrados, alagamentos, etc.) geralmente impacta temporariamente as condições do tráfego da via: velocidade, densidade e volume do tráfego. A inclusão dos incidentes nos modelos de simulação aumenta a confiabilidade dos resultados em casos atípicos de tráfego, quando do uso de DTA (SLOBODEN et al., 2012).

Incidentes podem ser confundidos com valores atípicos e vice-versa, sendo necessário efetuar a Análise dos Dados buscando separar os dados típicos dos dados de incidentes e dos valores atípicos dando o correto tratamento a cada um deles.

Em pesquisa inicial sobre a detecção automática de incidentes foram identificadas diversas possibilidades de abordagem as quais, por não serem objetivo direto desta tese, estão apresentados no Apêndice E

3.4.1 Valores atípicos - outliers

Os estudos de valores atípicos “outliers” já perfazem mais de 50 anos e diversas abordagens podem ser adotadas (GUO; HUANG; WILLIAMS, 2015). Comparação de 5 métodos de detecção de outliers: Grubbs, Dixon, Hampel, Quartil e GESD indicou que os 3 últimos são capazes de identificar diversos outliers numa única aplicação (KANNAN, 2013).

O teste GESD tem a mesma base do teste de Grubbs, tendo sido generalizado por Rosner (1983) para a detecção de vários valores atípicos num único procedimento.

Saha et al. (2016) utilizaram PCA e o Nível de Serviço da via para detecção e remoção de outliers em seu modelo de análise de tráfego, sendo que o nível de serviço foi verificado com base na quantidade de veículos lentos e na porcentagem de redução de velocidade. O teste T^2 de Hotelling foi utilizado em comparação ao PCA no processo de limpeza dos dados e o processo de limpeza foi feito de modo iterativo constatando-se que os novos dados se tornaram mais confiáveis. Neste caso o estudo verificou a necessidade de proceder a limpeza de modo iterativo, pois boa parte dos dados atípicos era relativo aos meios não motorizados (bicicletas, etc.). Destaca-se a abordagem utilizada nesse estudo, pois, além da análise estatística, incorpora o nível de serviço aproximando-se das técnicas de estudo de tráfego mais tradicionais.

3.4.2 Dados ausentes

Um tipo de falha que pode impactar os dados de séries temporais é a ausência de informação de uma ou mais fontes de dados (sensores, veículos, etc.) por um ou mais intervalos de medição.

Duas ações podem ser tomadas: exclusão dos intervalos sem dados, possível em algumas metodologias como no caso dos Dados em Painel, ou a complementação dos dados, como foi necessária na aplicação dos modelos STARIMA.

A complementação dos dados pode ser feita com base em valores estimados ou previstos por sistemas de estimação e previsão. Todavia, quando esses sistemas ainda não foram preparados para efetuar estimação e previsão de valores complementares, outras estratégias podem ser utilizadas, como por exemplo: interpolação de dados; uso de valores médios históricos; uso do último dado disponível.

Os modelos baseados em STARIMA necessitam de séries completas, tendo em vista sua abordagem que trabalha utilizando diretamente os dados dos intervalos anteriores. Assim, nesta tese, utilizamos a estratégia de repetir o último valor medido.

Para os estudos com Dados em Painel, não foi feita a complementação dos dados, tendo sido excluídos da série registros com os valores vazios, pois esse tipo de abordagem permite o uso de séries descontínuas e não balanceadas.

3.5 MEDIDAS DE DESEMPENHO E ACURÁCIA

Nos estudos de previsão de tráfego o Erro Percentual Absoluto Médio (MAPE) e a Raiz do Erro Quadrático Médio (RMSE) têm sido os mais aplicados (LAÑA et al., 2018; PAVLYUK, 2019). E, daí foram os escolhidos para a avaliação dos cenários propostos nesta tese.

As Expressões 3-9, 3-10, apresentam respectivamente os medidores de acurácia MAPE e RMSE, sendo: O_i o valor Observado, P_i o valor Previsto e n o tamanho da amostra. Nas duas comparações quanto menor o valor, melhor a acurácia do cenário.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{O_i - P_i}{O_i} \right| \quad (3-9)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (3-10)$$

Nos estudos relativos ao modelo STARIMA, o cenário adotado para cada um dos tipos de matriz testado é escolhido com base nos valores de acurácia (MAPE e RMSE) do período de treino. Em alguns casos, onde os resultados de dois cenários obtidos na etapa de treino são muito semelhantes, a escolha foi feita com base na avaliação desses cenários com os dados de teste.

No caso dos modelos de regressão com dados em painel, na etapa de treino, a comparação entre os cenários é feita diretamente com base no valor de verossimilhança e no Critério de Informação de Akaike (AIC) - Expressão 3-11. Quando da análise dos modelos com uso dos dados de teste, a comparação se dá através dos medidores de acurácia, MAPE e RMSE.

$$AIC = 2k - 2\ln(\hat{L}) \quad (3-11)$$

Sendo k o número de parâmetros utilizados no modelo e \hat{L} o máximo valor da função de verossimilhança.

Para comparação entre os modelos de estimação, dos dois métodos avaliados, serão utilizados os medidores de acurácia MAPE e RMSE, avaliando a acurácia obtida no modelo com relação aos valores estimados e os valores observados na base de testes.

3.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO 3

Neste capítulo foi apresentada a metodologia utilizada nesta tese, descrevendo-se cada um dos módulos da metodologia.

Os modelos estatísticos que serão utilizados no processo de estimação foram apresentados, tendo sido escolhidos por permitirem melhor entendimento dos fenômenos envolvidos, pela oportunidade em se contribuir significativamente com o aumento da base de conhecimento relacionada a eles, por permitirem a análise espaço-temporal e por serem escaláveis.

O preparo das matrizes de peso espacial foi detalhado, tendo sido apresentada a preparação para o uso inédito de matrizes agrupadas para a previsão e estimação de tráfego com uso do STARIMA. O teste comparativo com diversos pesos simétricos e assimétricos está descrito no capítulo 5.

Foi destacado o processo de pré-tratamento dos dados, uma vez que pode não ser trivial, dependendo do formato em que os dados são recebidos. A anonimização dos dados possibilitará sua utilização nesta e em outras pesquisas.

Os índices que serão utilizados para comparar os modelos e medir sua acurácia foram apresentados.

4 PREPARAÇÃO DOS DADOS DA PESQUISA

Para a aplicação e avaliação dos modelos propostos na metodologia escolhemos 3 dos principais corredores de tráfego do município de São Paulo para os quais estavam disponíveis os dados de sensores, do tipo radar, da SMMT-SP.

Os dados foram preparados de modo a possibilitar sua aplicação na etapa de estimação, descrita na metodologia.

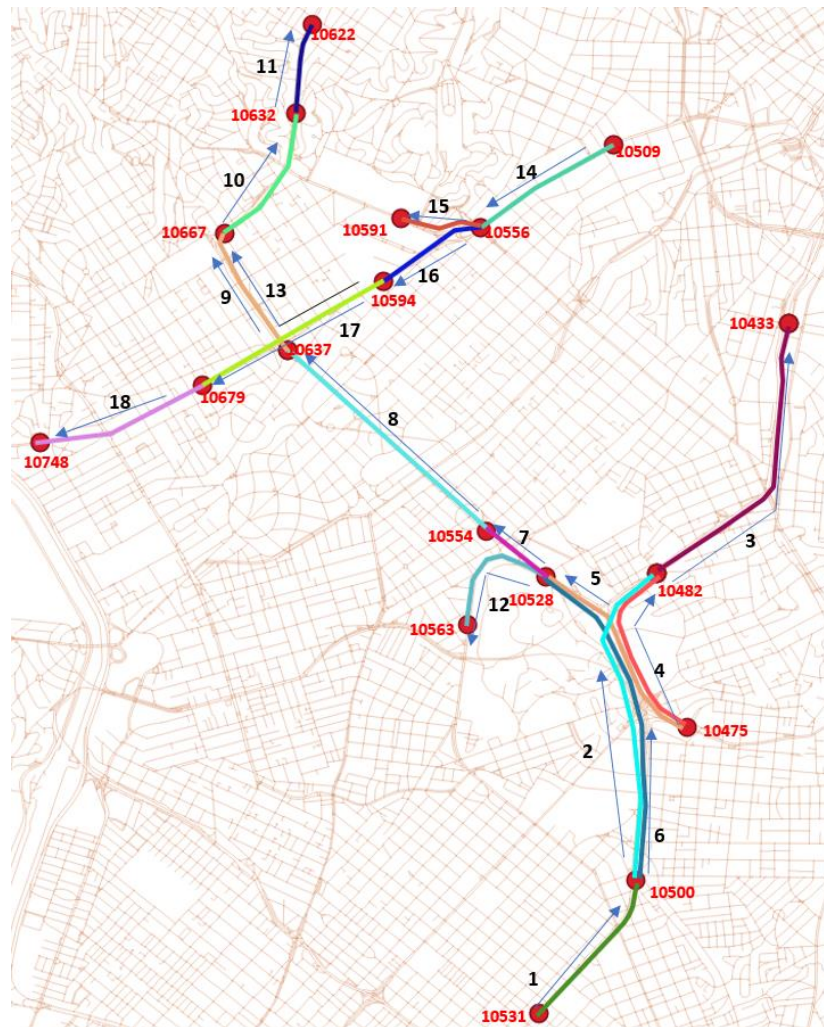
Neste capítulo apresentamos o local dos estudos e o tratamento de dados que precede a aplicação do módulo de estimação das características do tráfego.

4.1 LOCAL DA APLICAÇÃO DA PESQUISA

Para esta pesquisa foram selecionados 3 corredores de tráfego do Município de São Paulo, os quais estão apresentados na Figura 4-1, com os trechos em linhas coloridas, numerados de 1 a 18 e os sensores numerados em vermelho, com números que vão de 10433 a 10748.

A escolha desses corredores se deu com base na existência de dados de sensores e com o objetivo de incluir 3 características distintas de infraestrutura, sendo um eixo de via expressa, outro de via arterial e o terceiro uma via arterial com faixa exclusiva de ônibus à esquerda.

Figura 4-1 - Mapa com a indicação dos radares e links do estudo



Fonte: o autor

Com base na Figura 4-1 é possível identificar os 18 trechos escolhidos, definidos com base na presença de 18 sensores, cujas características principais estão apresentadas na Tabela 4-1.

Destaca-se que a identificação de eixo aqui apresentada foi utilizada durante os estudos com o STARIMA e que após sua conclusão constatou-se a necessidade de dividir o Eixo Brasil/Sumaré em 2. Assim, no estudo com Dados em Painel os trechos 7, 8 e 9 formaram o Eixo Brasil e os trechos 10 e 11 o Eixo Sumaré.

Tabela 4-1 – Características dos Trechos do estudo

Trecho	Eixo	Sensor		Extensão (km)	Período para obtenção do tempo de percurso (min)	Velocidade de Percurso		Tempo de Percurso	
		Início	Fim			Média	Desvio Padrão	Média	Desvio Padrão
1	23 de Maio	10531	10500	1,367	24	39,34	5,68	2,08	0,30
2	23 de Maio	10500	10482	2,756	44	25,47	3,83	6,49	0,98
3	23 de Maio	10482	10433	2,489	40	39,56	5,83	3,78	0,56
4	*	10475	10482	1,688	28	27,23	2,47	3,72	0,34
5	*	10475	10528	1,733	28	28,74	4,38	3,62	0,55
6	*	10500	10528	2,828	44	23,46	2,42	7,23	0,75
7	Brasil/Sumaré	10528	10554	0,601	12	21,67	3,45	1,66	0,26
8	Brasil/Sumaré	10554	10637	2,128	32	13,07	0,95	9,77	0,71
9	Brasil/Sumaré	10637	10667	1,170	20	17,00	2,03	4,13	0,49
10	Brasil/Sumaré	10667	10632	1,185	20	30,65	2,93	2,32	0,22
11	Brasil/Sumaré	10632	10622	0,748	12	21,73	2,16	2,07	0,21
12	*	10528	10563	1,179	20	19,18	0,88	3,69	0,17
13	*	10594	10667	2,005	32	9,59	3,75	12,54	4,91
14	Rebouças/Cons	10509	10556	1,228	20	16,06	1,51	4,59	0,43
15	*	10556	10591	0,658	12	25,60	3,10	1,54	0,19
16	Rebouças/Cons	10556	10594	0,873	16	23,23	2,52	2,25	0,24
17	Rebouças/Cons	10594	10679	1,633	24	11,53	1,31	8,50	0,97
18	Rebouças/Cons	10679	10748	1,355	20	11,51	1,94	7,06	1,19

*trechos que se conectam aos Eixos estudados

** Os trechos da Brasil/Sumaré foram separados para o estudo com Painel de Dados

Fonte: o autor

Cabe indicar que, embora a rede esteja representada por 18 trechos de estudo, na verdade cada trecho poderia ser decomposto em diversos subtrechos, definidos pela existência de cruzamentos e semáforos ao longo deles. No entanto, não estavam disponíveis informações para cada subtrecho e para cada via que entrecorta as vias analisadas, não sendo possível precisar a quantidade de veículos que acessam e saem da rede em cada uma dessas vias não monitoradas.

4.2 DADOS DE SENSORES

Nesta pesquisa os dados dos diversos equipamentos instalados numa mesma seção da via foram agrupados, a fim de caracterizar um único “sensor” por ponto de fiscalização. Isto não implica numa regra universal, sendo possível que em outras pesquisas cada equipamento seja considerado como um sensor separado. Como exemplo temos os estudos em andamento para utilização de Dados em Painel para complemento de dados de volume (fluxo) de veículos em conjunto com a equipe de planejamento da CET, conforme apresentado no Apêndice B.

Os dados de sensores utilizados nesta pesquisa foram obtidos junto à SMMT-SP e são relativos ao ano de 2018. Naquele ano o Município de São Paulo contava com 1630 equipamentos que fiscalizavam 1005 pontos diferentes da cidade. Os equipamentos eram locados, sendo que a cidade estava dividida em quatro áreas e cada área correspondia a um “lote”. Cada lote foi contratado de um fornecedor diferente. Os sensores utilizados nesta pesquisa fazem parte da área do lote 3.

Os sensores são do tipo radar de fiscalização de velocidade, sendo que alguns deles também fiscalizam infrações relativas à semáforo vermelho, faixas exclusivas de ônibus e conversão proibida. Estavam disponíveis 4 tipos de sensores, de acordo com o tipo de equipamento instalado:

- a) radar portátil: equipamento portátil utilizado por agentes da CET-SP em pontos pré-determinados, em sistema de rodízio;
- b) radar estático: equipamento móvel que é instalado em tripés ou outros tipos de suporte. Normalmente utilizado para a fiscalização de uma ou duas faixas de tráfego;
- c) radar fixo: equipamento instalado em infraestrutura fixa, ao lado ou sobre a via, que utiliza laços detectores magnéticos;
- d) lombada eletrônica: equipamento fixo, com totem e painel de mensagem variável, que apresenta a velocidade medida em tempo real.

A implantação de equipamentos fixos (radar fixo e lombada eletrônica) pode envolver mais de um equipamento na mesma seção da via, de modo que, nesta pesquisa o conjunto de equipamentos foi considerado como um único sensor. Apenas equipamentos fixos foram inseridos na pesquisa, pois isso possibilitava a análise de dados de cada ponto de fiscalização ao longo de diversos dias.

4.2.1 Dados disponíveis

Cada equipamento envia um arquivo de texto com os dados coletados em intervalos de 4 minutos. Assim, mais de 570 mil arquivos eram gerados diariamente, com os dados descritos na Tabela 4-2.

Tabela 4-2 – Dados originais dos sensores do tipo radar da SMMT-SP

Nome do Campo	Observações
Código da empresa	L1, L1, L3 e L4
Data	Data/Hora no formato: AAAAMMDD HHMMSS
Local	Número ID de um equipamento. Pode existir um equipamento ou mais na mesma seção de via. A informação de faixa e entre-faixa permite saber a posição exata. Para obter o volume numa seção de via é necessário somar os vários radares da mesma seção.
Faixa	Faixa de Rolamento numerada da esquerda para a direita
Entre-Faixa	Quando disponível equipamento específico para fiscalização de motos, entre as faixas de veículos
Registro	ID do Registro daquela passagem, naquele equipamento e naquele mês.
Tipo de registro	Tipo do registro: 0 = comum = registro de veículo não infrator; 1 = infrator = registro de veículo infrator; 2 = s/registo = registro em branco em caso de não passagem de veículo sendo, obrigatórios, neste caso, os campos de 1 a 7, sendo que os demais campos não devem ser informados.
Placa	Placa do veículo no formato NNNNNN0
Espécie	Espécie do veículo: moto = 0, passeio = 1, ônibus = 2 ou caminhão = 3
Classe	Classificação de tipo de veículo: leve = 0 ou pesado = 1.
Comprimento	Comprimento medido do veículo - Em decímetros (dm)
Velocidade Pontual	Velocidade Pontual: Em decímetros por segundo (dm/s)
Tempo de ocupação	Tempo de ocupação na área de detecção: Em milissegundos (ms)

Fonte: adaptado de (SECRETARIA MUNICIPAL DE TRANSPORTES, 2019)

Embora os dados sejam enviados agrupados, é possível verificar as informações individuais de cada veículo registrado. Assim, é possível identificar a passagem de um mesmo veículo em mais de um local viabilizando, por exemplo, seu uso na definição de tempos de percurso e na elaboração de matrizes Origem Destino (OD). Nesta tese efetuamos seu uso na determinação dos tempos de percurso.

4.2.2 Dados utilizados

Após a aplicação do script de anonimização e organização dos dados, descrito no item 4.3, foi efetuada a seleção dos registros relativos aos 18 sensores incluídos no estudo, gerando-se arquivos que continham apenas esses dados para cada dia. Depois de produzidas as séries temporais foram gerados arquivos que continham todos os dias escolhidos para o estudo.

Para os estudos com STARIMA, realizados na primeira etapa da pesquisa, foram utilizados apenas os dados dos dias úteis do mês de março de 2018. Este mês foi escolhido pela presença de dias típicos e atípicos para o tráfego. Os dias atípicos correspondiam a dias com fortes chuvas e feriados, como por exemplo o dia 20 de março, que foi uma terça-feira com tráfego atípico, decorrente de fortes chuvas no período da tarde, implicando inclusive na interdição da Av. 23 de Maio e o dia 30 de março, que foi o feriado da “sexta-feira santa”. Além disso, como forma de verificar a robustez do modelo, no caso do STARIMA, foi incluído - na base de teste - o dia 06/09, uma quinta-feira, véspera do feriado de independência do Brasil e que está 6 meses à frente dos dias de treino.

Para a aplicação com STARIMA foram escolhidos os dias 01, 06, 08 e 13/03 até às 14h00, duas terças e duas quintas-feiras, para treino do modelo. Para a base de testes foram utilizados os dias 13/03, à partir das 15h00, 20 e 21/03 e 06/09.

Todos os dados foram utilizados no estudo preliminar de matrizes de relação espacial, efetuado com os dados de volume (fluxo).

Para a análise comparativa com os modelos efetuados com a técnica de Dados em Painel, os dados de velocidade foram tratados na mesma base de treino, mas apenas os dias 20 e 21/03 foram utilizados na base de testes.

Nos estudos com Dados em Painel, concluídos na segunda fase da pesquisa, todos os 31 dias do mesmo mês de março foram utilizados, o que possibilitou considerar as diferenças sazonais que ocorrem ao longo da semana. Os 19 primeiros dias do mês serviram como base de treino e os últimos 12, que incluem os dias atípicos já citados, para teste. Na análise comparativa com o modelo STARIMA, apenas os dias 20 e 21/03 foram utilizados.

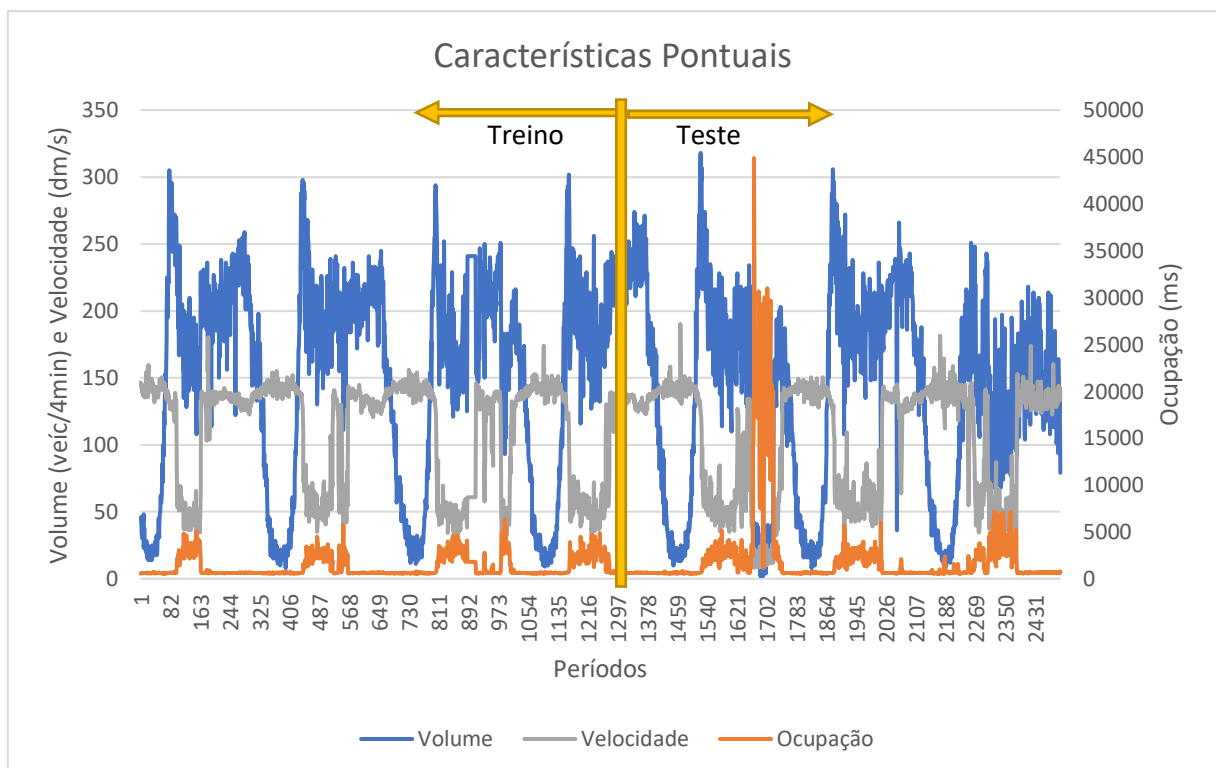
4.2.3 Visão preliminar das características do tráfego

Com base nos dados disponíveis foi possível obter diversas características do tráfego nas vias avaliadas.

Como exemplo, na Figura 4-2 estão apresentadas as características do tráfego, velocidade, fluxo (volume) e ocupação, num dos sensores da Av. 23 de Maio, sentido centro, durante uma semana, iniciando-se pela segunda-feira.

Verifica-se que as 3 características possuem um perfil sazonal, sendo que o volume é reduzido nas madrugadas e mantém-se alto no restante do dia. Ocupação e Velocidade pontual possuem certa correlação inversa, quando uma sobe a outra tende a se reduzir, dessa forma, a velocidade na hora do pico se reduz significativamente e a ocupação média aumenta, ambas em decorrência do aumento da densidade do fluxo.

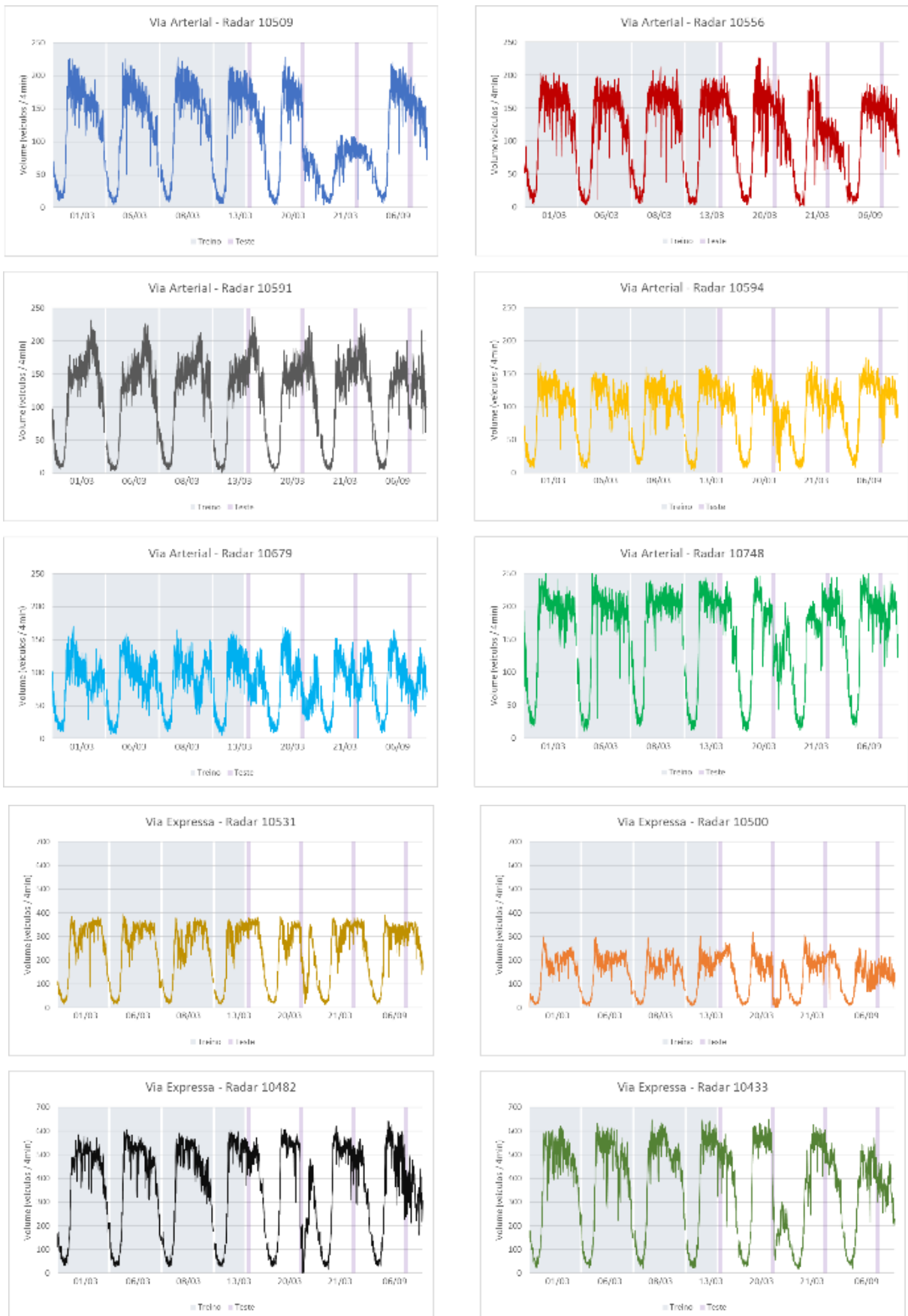
Figura 4-2 – Características do Tráfego



Fonte: o autor

A distribuição do fluxo, a cada 4 minutos, nos 10 sensores dos eixos Rebouças (via arterial) e 23 de Maio (via expressa) é apresentada na Figura 4-3, considerando apenas os dias utilizados no estudo com STARIMA. É possível verificar que cada sensor apresenta volumes diferentes, embora mantendo as características sazonais.

Figura 4-3 - Distribuição do Fluxo em 10 radares nos dos dias de treino e teste do modelo STARIMA



Obs. Os dias de treino estão marcados em cinza e os períodos de teste em lilás.

Fonte: o autor

Também é possível verificar a existência de dias atípicos, como o 5º dia dessa série, que corresponde ao dia 20/03, onde ocorreram fortes chuvas no município. Destaca-se que alguns sensores apontam maior impacto desse fato, em especial o sensor 10500, localizado logo após um dos pontos onde ocorreu o bloqueio da via, por conta de alagamento.

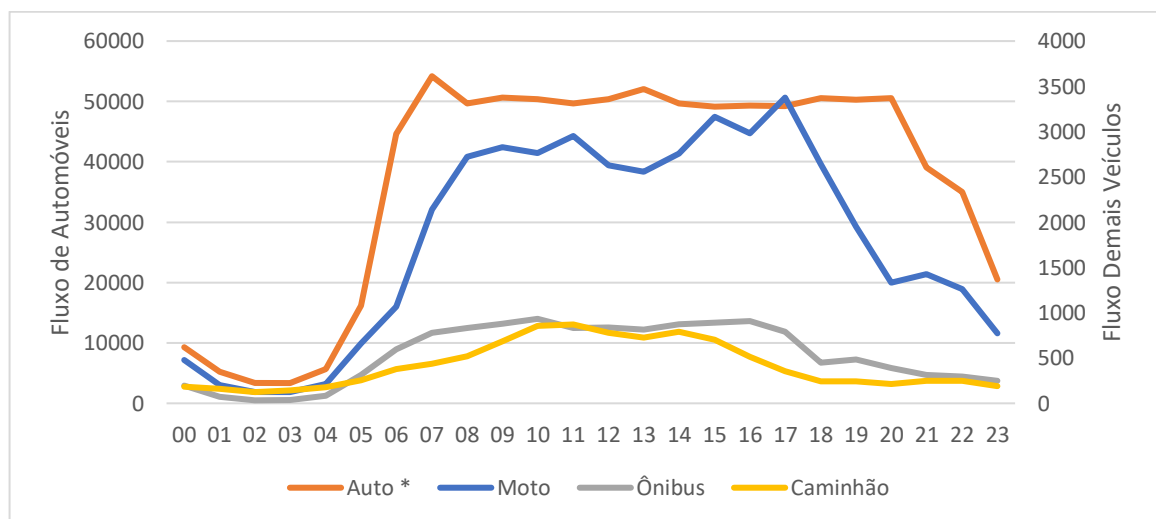
No 6º dia (21/03) se verifica nos sensores 10509 e 10556 fluxos muito abaixo do normal e com variação diferente dos demais sensores. Isto pode decorrer de falha, em parte dos equipamentos que compõem esses sensores, ou de alguma interferência no tráfego da Rua da Consolação, onde esses equipamentos estão instalados.

Como ocorre grande flutuação de fluxo, ao longo do dia, e sendo o modelo dependente dessa flutuação, torna-se adequada a elaboração de modelos específicos para cada período característico ao longo do dia, o que possibilita maior acurácia na previsão (DUAN et al., 2019). Após comparação do volume e da velocidade dos diversos pontos de contagem, foi efetuada a divisão do dia em 6 períodos característicos indicados na Tabela 4-4.

4.2.4 Identificação dos veículos

Um dia típico apresenta em torno de 1 milhão de registros, considerando apenas os 18 sensores incluídos na pesquisa.

Figura 4-4 – Fluxo horário por tipo de veículo para um dia típico



Obs. A escala da esquerda corresponde ao fluxo de automóveis e a da direita aos demais veículos

Fonte: o autor

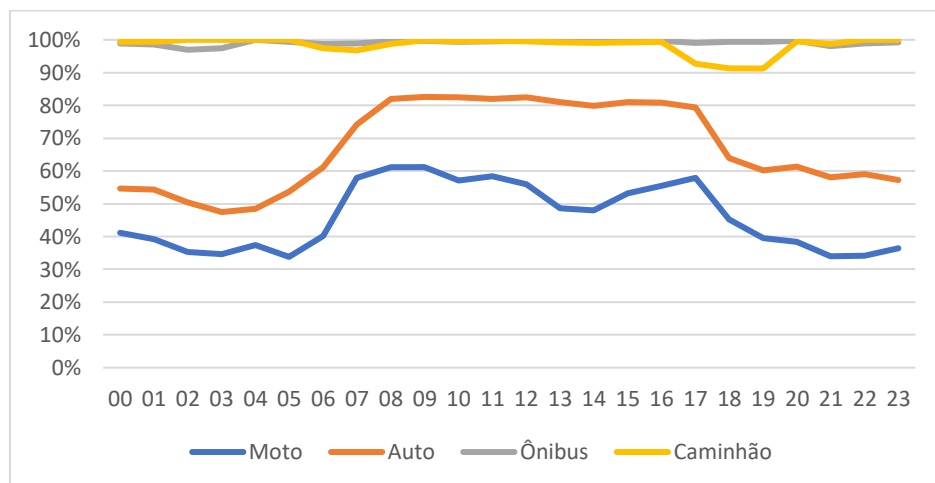
A composição do fluxo se caracteriza pela detecção de 93% de automóveis, 4% de motocicletas, 2% de ônibus e 1% de caminhões. A distribuição do fluxo ao longo das horas do

dia, classificada por tipo de veículo, está apresentada na Figura 4-4. Verifica-se que a madrugada possui menor fluxo para todos os tipos de veículo, e que o fluxo de motocicletas e ônibus começa a decair no início do pico da tarde. Com relação ao fluxo de caminhões cabe indicar que existe restrição a esse tipo de veículo nos horários de pico, de modo que a maior concentração ocorre entre 10h e 16h.

No preparo inicial dos dados de sensores foi verificada a presença de períodos sem dados, em alguns equipamentos, e a indisponibilidade de dados, com a placa dos veículos, em 28% dos registros, o que poderia impactar na aplicação do script relativo ao tempo de percurso nos trechos. Essa perda de informação é mais significativa para as motocicletas e automóveis no período noturno.

O gráfico da Figura 4-5 apresenta os dados de um dia típico de coleta de dados, sendo verificado que, para os ônibus e caminhões a identificação supera os 90% em qualquer horário. Para os automóveis, ao longo do dia ocorre a identificação de mais de 80% dos veículos, todavia à noite esse índice cai para 60% e atinge até menos de 50% nos horários de menor fluxo. As motos são identificadas em menos de 60% dos casos, durante o dia, e à noite a porcentagem de identificação cai abaixo de 40%.

Figura 4-5 – Porcentagem de veículos identificados ao longo do dia, de acordo com o tipo e veículo



Fonte: o autor

O pareamento de placas em sensores consecutivos foi utilizado para a definição do tempo de percurso dos trechos, sendo que, em cerca de 26% dos intervalos não foi possível obter o valor, devido à impossibilidade de identificar ao menos um veículo nos sensores, de início e de final do trecho. Isto ocorre por quatro motivos:

- a. Falha de um sensor ou na transmissão do dado via rede, que entre 06h e 23h ocorreu em 1,5% dos casos
- b. Ausência da passagem de veículos pelo sensor, decorrente de bloqueio da via, como o que ocorreu em 20/03 devido às chuvas intensas, resultando em 30% de falhas durante o período de chuvas.
- c. Ausência de passagem de veículos durante o intervalo de medição, como ocorre frequentemente nas madrugadas, entre 01h e 05h
- d. Baixo nível de identificação, decorrente de chuvas e iluminação (período noturno), associado à redução do fluxo que ocorre entre 23h e 06h

4.2.5 Dados de chuva

Os dados de chuvas foram fornecidos pelo Sistema de Alerta a Inundações de São Paulo (SAISP). São dados de pluviometria medidos a cada 10 minutos, para cada uma das 31 estações meteorológicas que cobrem o Município de São Paulo. Os dados recebidos estavam organizados com base na quantidade de precipitação, em cada uma das 33 Subprefeituras do Município.

Para uso no sistema foi necessário definir uma relação entre sensores e estações meteorológicas, o que foi feito cruzando-se a posição georreferenciada dos mesmos com a utilização do software QGIS

A definição foi feita com base em polígonos de Voronoi, criados em torno das estações meteorológicas. Todos os sensores que estavam no polígono de uma estação foram vinculados a ela.

Os dados foram tratados de modo a estabelecer o índice pluviométrico acumulado da última hora, sendo obtidos valores para cada estação meteorológica. Posteriormente os dados de chuva foram inseridos nos bancos de dados das séries temporais, possibilitando seu uso nos estudos multivariados, como nas regressões com Dados em Painel.

4.3 PRÉ-PROCESSAMENTO DOS DADOS

Quando da recepção dos dados dos sensores, existe a necessidade de efetuar um tratamento prévio, preparando os dados para serem utilizados nos diversos passos da metodologia proposta.

A etapa de pré-processamento dos dados envolve receber, abrir, ler, analisar e tratar os dados recebidos. Neste caso, o tratamento consiste em eliminar os registros corrompidos e valores extremamente atípicos e formatar os dados dentro do padrão adequado, para uso nas demais etapas da metodologia.

Os arquivos de dados dos sensores foram fornecidos pela SMMT-SP, compactados de acordo com a data, lote e equipamento. Os dados originais continham a informação da placa dos veículos. Esse dado é considerado sensível, pois permite, de modo objetivo, efetuar buscas de um veículo específico.

O script para efetuar a descompactação e agrupamento automático, dos arquivos, foi o primeiro a ser elaborado em Python para esta pesquisa. Ele gerava arquivos com extensão “csv”, com dados diários e com todos os registros efetuados.

O passo seguinte envolveu a necessidade de anonimizar os dados, cumprindo uma exigência efetuada pela SMMT-SP, quando da liberação de acesso. Neste caso, cabe a observação de que esses dados, não anonimizados, estavam indisponíveis para uso, tanto de outros pesquisadores como de órgãos públicos, como a CET-SP, a SPTRans. Assim, o processo de anonimização dos dados possibilitou a utilização deles por outros pesquisadores.

O processo de anonimização foi efetuado de modo a substituir a placa por um ID, destruindo-se a informação original, a fim de impossibilitar o acesso. O ID atribuído é válido para apenas um dia, de modo que é possível elaborar estudos de Origem e Destino, de trajetos e de tempos de percurso, desde que limitados às 24 horas do dia (ITABORAI, 2022).

Ao mesmo tempo em que foi feita a anonimização dos dados, a coluna de velocidade média foi excluída, pois não continha dados e foi inserida uma coluna com o número final da placa, de forma a permitir análises futuras relativas ao rodízio municipal (ITABORAI, 2022).

As ocupações negativas, por possuírem característica atípica, foram excluídas.

4.4 MONTAGEM DAS SÉRIES TEMPORAIS

Os dados dos sensores foram tratados de modo a agrupá-los em períodos de 4 minutos, uma vez que é este o intervalo de envio de dados pelos equipamentos de campo. O agrupamento foi feito para os registros de cada sensor em separado, possibilitando a análise por sensor, ou eixo de tráfego.

4.4.1 Dados pontuais

Com base nos dados pré-processados para cada sensor, foram obtidos os valores correspondentes a cada intervalo de 4 minutos, sendo gerado um arquivo csv para cada dia:

O procedimento de preparo dos dados de radares em intervalos acompanhou o seguinte procedimento:

- a) Separar, num arquivo csv, os dados diários dos sensores escolhidos no estudo;
- b) considerar todos os veículos que passarem pelo sensor “L” durante o intervalo j , a cada intervalo de 4 minutos, independentemente de terem sido identificados.
- c) calcular:
 - velocidade média pontual (VL_j) = média das velocidades medidas durante o intervalo j no sensor L ;
 - ocupação média pontual (OL_j) = média das ocupações medidas durante o intervalo j ;
 - ocupação no intervalo = (OC_j) soma de todas as ocupações ocorridas durante o intervalo j no sensor L .
 - volume (Q_L^j) = contagem dos veículos i que passaram no sensor L no intervalo j ;
 - volume equivalente (EL_j) = contagem classificada e ponderada dos veículos i que passaram no sensor L no intervalo j . Para isto deve ser feita a contagem classificada dos veículos por espécie no intervalo (Q_{Lej}) e multiplicar o volume por espécie pelo fator de equivalência F_e definidos na Tabela 4-3, considerando-se todos os caminhões como tendo 2 eixos. Calcular o volume equivalente E_R^j com base na Expressão 4-1 ;

Tabela 4-3 - Fatores de equivalência por UCP

TIPO	FATOR DE EQUIVALÊNCIA
Automóvel	1,00
Moto	0,33
Ônibus	2,00
Caminhão de 2 eixos	2,00
Caminhão de 3 eixos	3,00

Fonte: (CONSELHO NACIONAL DE TRÂNSITO CONTRAN, 2014)

$$E_R^j = Q_{Re}^j \cdot F_e \quad (4-1)$$

- d) Montar um arquivo csv para cada dia de estudo, com os valores obtidos no item anterior, contendo os valores de cada intervalo “j”;
- e) Unificar os arquivos de cada variável, colocando-se em sequência todos os dias de treino e teste num único arquivo.

Observa-se que podem ser obtidas séries separadas por tipo de veículo, todavia verificou-se uma grande quantidade de valores vazios e volumes unitários, motivo pelo qual foram mantidos apenas os dados agregados.

Os dados foram organizados em dois formatos, de acordo com a aplicação pretendida:

- a) Para a aplicação com o modelo STARIMA os dados foram organizados de modo a obter uma tabela com a série temporal de cada sensor e colocada numa coluna. Foram elaborados 3 bancos de dados, sendo um para volume, outro para velocidade média e outro para ocupação no intervalo. A Figura 4-6 apresenta uma amostra dos dados de volume;

Figura 4-6 – Exemplo da série temporal pontual dos volumes para uso com STARIMA

Data	10531	10500	10482	10433	10475	10528	10554	10637	10667	10632	10622	10563	10509	10556	10591	10594	10679	10748
2018-02-28 23:56:00	97	6	167	180	34	44	30	58	62	65	56	36	72	52	97	65	101	158
2018-03-01 00:00:00	95	3	111	151	22	39	19	43	53	66	79	33	47	52	93	62	97	140
2018-03-01 00:04:00	111	19	127	174	32	70	33	80	63	62	61	26	53	59	73	57	86	160
2018-03-01 00:08:00	98	48	173	144	23	69	24	42	42	60	69	38	70	66	52	57	48	139
2018-03-01 00:12:00	88	59	141	166	28	61	27	57	55	54	52	38	49	66	80	71	78	79
2018-03-01 00:16:00	93	58	139	148	22	65	36	71	62	49	50	20	53	65	57	52	52	121
2018-03-01 00:20:00	85	53	132	178	23	38	14	64	68	54	48	38	55	59	52	49	49	134

Fonte: do autor

- b) No caso do estudo com o modelo de Dados em Painel foi criada uma série multivariada, com uma coluna para cada variável, e colunas onde se indicaram o intervalo de medição, o sensor e o eixo de estudo, permitindo a elaboração de modelos hierárquicos com repetição nas medições. Foram ainda incluídas as identificações dos sensores e trechos a montante e jusante de cada sensor,

possibilitando análises desse relacionamento em modelos baseados em Dados em Painel Longitudinal e Hierárquicos.

O banco de dados obtido tinha então 6 colunas de identificação do local e de seus vizinhos por contiguidade e outras 35 colunas com informações que foram utilizadas nos modelos, compondo, assim, as 41 colunas descritas na Tabela 4-4.

Tabela 4-4 – Características dos dados pontuais

Variável	Nome da Variável no BD	Descrição	Observações
Eixo Viário ¹	Eixo	Todos os trechos da mesma via no mesmo sentido	Indicação do Eixo viário. Nas regressões são caracterizados como classes de 0 a 3
Sensor ¹	Num_agrupado	Inclui os dados de todos os equipamentos da mesma seção	Variável categórica de identificação do sensor
Sensor a montante ¹	radar_montante	Identifica o Sensor a montante	Variável categórica (Num_Agrupado) do sensor a montante
Sensor a jusante ¹	radar_jusante	Identifica o Sensor a jusante	Variável categórica (Num_Agrupado) do sensor a jusante
Trecho a montante ¹	trecho_montante	Identifica o Sensor a montante	Variável categórica de identificação do trecho a montante
Trecho a jusante ¹	trecho_jusante	Identifica o Trecho a jusante	Variável categórica de identificação do trecho a jusante
Número de Faixas ¹	Num_faixa	Quantidade de faixas de tráfego	Variável categórica que indica a quantidade de faixas. 0 = 3 faixas, 1=4 faixas, 2 = 5 faixas e 3 = 6 faixas
Faixa de Ônibus ¹	Fx_Onibus	Identifica a presença de faixa exclusiva de ônibus	Variável categórica com valor 0 na ausência de faixa de ônibus e 1 na presença
Classe da Via ¹	Classe	Indica a classe da via	Variável categórica com valor 0 para vias arteriais e 1 para a via expressa
Estação Meteorológica ¹	Est_Meteorol	Estação meteorológica que corresponde ao sensor	Dado categórico utilizado no vínculo entre dados meteorológicos e o sensor
Chuva	CH_1h	Índice pluviométrico acumulado em 1 h	Em mm, índice acumulado na estação meteorológica na última 1 hora. Foram criadas duas variáveis dependendo do tipo de uso.
Data	Data	Data/hora da amostra no formato data/hora	Mantida como base para emissão de relatórios e imagens
	data3	Data/hora da amostra no formato numérico. Indica o número do intervalo	Datas transformadas para número sequencial a fim de possibilitar seu uso no modelo. Cada intervalo de 4 minutos corresponde a 1 unidade de tempo. Foram utilizadas 3 variáveis de acordo com a função adotada
	data4		
	data 5		
Período do dia ¹	Periodo	Período do dia	Períodos característicos do fluxo na Cidade: 0h-04h; 04h-06h; 06h-10h; 10h-16h; 16h-20h; 20-24h
Dia da Semana ¹	dia_sem	Dia da Semana	Indica o dia da semana, sendo 0 igual a segunda-feira

¹ variáveis categóricas

Tabela 4-4 - Continuação

Variável	Nome da Variável no BD	Descrição	Observações
Dia útil ¹	dia_u	Dia útil	Indica se é dia útil ou final de semana, sendo 0 para dia útil
Velocidade Média	Vel_media	Media das velocidades detectadas no grupo no intervalo	Em decímetros por segundo. Para manter a compatibilidade com os demais campos. Transformado para km/h para facilitar o entendimento e análise
Desvio Padrão da Velocidade	Vel_std	Desvio Padrão da velocidade no intervalo	Desvio padrão da velocidade medido durante o intervalo
Velocidade Mediana	Vel_mediana	Mediana das velocidades detectadas no grupo no intervalo	Em decímetros por segundo. Esta unidade permite o uso com números inteiros reduzindo a demanda de processamento. Transformado para km/h para facilitar o entendimento e análise
Velocidade Regulamentada	Vel_Reg	Velocidade Regulamentada no trecho ou sensor	Em decímetros por segundo. Ocorreu com 2 valores 139 (50km/h) e 167 (60km/h)
50% da Velocidade Regulamentada	vel_reg_50	Metade da Velocidade Regulamentada	Utilizada na definição do tipo de velocidade
Velocidade Tipo ¹	vel_tp	Classificação em relação à velocidade regulamentada	0 para velocidades entre 50% e 100% da velocidade regulamentada e 1 se menor que 50%
Ocupação Média	Ocup_media	Ocupação média no intervalo	Em milésimos de segundo, indica o tempo médio de permanência dos veículos na área de detecção durante o intervalo de 4 min
Desvio Padrão da Ocupação	Ocup_std	Desvio Padrão da ocupação no intervalo	Desvio padrão da ocupação medido durante o intervalo
Ocupação Mediana	Ocup_mediana	Mediana da Ocupação no intervalo	Em milésimos de segundo, indica a mediana dos tempos de permanência dos veículos na área de detecção durante o intervalo de 4 min
Log da Ocupação Média	Log_ocup_media	Logaritmo da ocupação média	Utilizado em alguns modelos, com Dados em Painel, em substituição da Ocupação Média
Log da Ocupação Mediana	Log_ocup_mediana	Logaritmo da ocupação média	Utilizado em alguns modelos, com dados em Painel, em substituição da Ocupação Média
Ocupação no intervalo	ocupancia	Ocupação do sensor durante o intervalo	Em milésimos de segundo, indica a soma das ocupações no intervalo de 4 min
Volume	Volume	Quantidade de veículos detectados no intervalo	Total de veículos de cada espécie detectada no intervalo
Volume Equivalente	Vol_eq	Volume em veículos equivalentes	Quantidade obtida pela ponderação do volume pelo tipo de veículo
Velocidade * Volume	vel_vol	Produto da velocidade média pelo volume	Variável multiplicativa utilizada em alguns modelos com Dados em Painel
Velocidade * Log Ocupação	vel_oclog	Produto da velocidade média pelo log da ocupação	Variável multiplicativa utilizada em alguns modelos com Dados em Painel
Log Ocupação * Volume	oclog_vol	Produto do log da ocupação pelo volume	Variável multiplicativa utilizada em alguns modelos com Dados em Painel

¹ variáveis categóricas

Tabela 4-4 - Continuação

Variável	Nome da Variável no BD	Descrição	Observações
Volume a montante	Vol_RM	Volume no sensor a montante do sensor avaliado	Utilizado em alguns modelos, com Dados em Painel com aspecto espaço-temporal
Volume a jusante	Vol_RJ	Volume no sensor a jusante do sensor avaliado	Utilizado em alguns modelos, com Dados em Painel com aspecto espaço-temporal
Velocidade mediana a montante	Vel_md_RM	Velocidade mediana no sensor a montante do sensor avaliado	Utilizado em alguns modelos, com Dados em Painel com aspecto espaço-temporal
Velocidade mediana a jusante	Vel_md_RJ	Velocidade mediana no sensor a jusante do sensor avaliado	Utilizado em alguns modelos, com Dados em Painel com aspecto espaço-temporal
Ocupação a montante	Ocup_md_RM	Ocupação mediana no sensor a montante do sensor avaliado	Utilizado em alguns modelos, com Dados em Painel com aspecto espaço-temporal
Ocupação a jusante	Ocup_md_RJ	Ocupação mediana no sensor a jusante do sensor avaliado	Utilizado em alguns modelos, com Dados em Painel com aspecto espaço-temporal
Inverso do Volume a montante	I_Vol_RM	Inverso do Volume no sensor a montante do sensor avaliado	Utilizado em alguns modelos, com Dados em Painel com aspecto espaço-temporal
Inverso do Volume a jusante	I_Vol_RJ	Inverso do Volume no sensor a jusante do sensor avaliado	Utilizado em alguns modelos, com Dados em Painel com aspecto espaço-temporal

¹ variáveis categóricas

Fonte: o autor

No caso das variáveis relativas aos sensores de jusante e montante, foi dado tratamento específico para os sensores de borda da rede. Neste caso, optou-se pela complementação dos dados através da criação de vizinhos fictícios fora da rede, visando manter a maior quantidade possível de registros, permitindo assim uma comparação mais direta entre os modelos com e sem vizinhos. As características dos sensores fictícios foram calculadas como a média dos valores entre o sensor de borda e o seu primeiro vizinho dentro da rede.

Por fim, para o tratamento com Dados em Painel, foram estabelecidas séries defasadas em até 8 intervalos, para as variáveis de velocidade média, volume, ocupação e ocupação no intervalo. Isso possibilitou que fosse verificado o impacto de medições passadas do sensor avaliado sobre si e dos sensores a montante e jusante sobre ele.

4.4.2 Dados de trecho

Os dados de Trecho consideraram que o fluxo de veículos é direcional e acompanha o traçado das vias, sendo possível utilizar o conceito de montante e jusante do fluxo. Assim, os

tempos de percurso do sensor “nm” até um sensor “nj” (par $nm\ nj$) no intervalo “j”, sendo “nm” um ponto à montante de “nj”, são denominados $T_{nm\ nj}^j$ e são definidos da seguinte forma:

- a) Pesquisar todos os veículos “i” que chegaram em “nj” no intervalo “j” e registrar o instante de chegada como sendo: t_{ij}^j = instante “t” da chegada do veículo “i” em “nj” durante o intervalo “j”;
- b) Pesquisar se os mesmos veículos “i” passaram por “nm” nos últimos “x” intervalos “j” e registrar o momento de sua última passagem “ t_{ix}^j ”.;
 - “x” foi obtido com base no tempo necessário para percorrer o trecho “nm nj” com velocidade de 4km/h, que corresponde a aproximadamente 1m/s. O limite aqui estabelecido visa restringir a pesquisa em intervalos anteriores, de modo a diminuir o esforço computacional e reduzir a possibilidade de identificação do mesmo veículo por mais de uma vez.
 - se houver mais de um registro do veículo “i” em “nm” apenas o registro mais recente era mantido;
- c) Calcular o tempo de percurso, em segundos, para todos os veículos “i” encontrados nos dois sensores, subtraindo-se o instante de chegada no sensor de fim de trecho do instante de passagem pelo sensor localizado no início do trecho: $Timj^j = t_{ij}^j - t_{inm}$ para o intervalo “j”, sendo: $Timj^j$ = tempo de percurso do veículo “i” entre os radares “nm” e “nj” no intervalo “j”. Valores negativos eram descartados;
- d) Fazer a mediana dos tempos de percurso $Timj^j$ obtidos para o par “nm nj” no intervalo “j”, determinando $T_{nm\ nj}^j$;
- e) Calcular a velocidade de percurso de cada par $V_{nm\ nj}^j = T_{nm\ nj}^j / D_{nm\ nj}$, onde $V_{nm\ nj}^j$ é a velocidade, em m/s, dos veículos no trecho “nm nj” para o intervalo “j” e $D_{nm\ nj}$ é a distância em metros entre os dois pontos, medida de acordo com o traçado da via.
- f) Contar a quantidade de veículos que foram detectados nos dois sensores e que, portanto, foram utilizados nos cálculos da velocidade e de tempo de percurso.

O tempo de percurso, entre sensores não vizinhos, foi obtido através da soma dos tempos de trajeto pelos trechos que compõem o percurso. A opção de uso da mediana se deu a fim de reduzir o impacto de outliers, reduzindo assim a necessidade de tratamentos subsequentes dos dados.

A velocidade de percurso $V_{nm nj}^j$ foi obtida dividindo-se a distância do trecho “ $nm nj$ ” pelo tempo de percurso $T_{nm nj}^j$. A velocidade de percurso entre radares não vizinhos foi obtida pela média harmônica das velocidades dos trechos.

Os dados de trechos a montante e jusante de cada sensor, obtidos nesta etapa, estão apresentados na Tabela 4-5. Estes dados foram diretamente utilizados nas análises com Dados em Painel. Para os modelos com STARIMA, os valores médios, do período de estudo, foram considerados nas montagens das matrizes de peso espacial apresentadas no item 3.3.

Tabela 4-5 – Características dos dados de trecho

Variável	Nome da Variável no BD	Descrição	Observações
Distância do sensor a montante	Dist_M	Distância do sensor a montante até o sensor avaliado	medida em km
Distância do sensor a jusante	Dist_J	Distância do sensor a jusante até o sensor avaliado	Medida em km
Tempo de Percurso a montante	Tempo_TM	Tempo de Percurso do sensor a montante até o sensor avaliado	Medido em segundos.
Tempo de Percurso a jusante	Tempo_TJ	Tempo de Percurso do sensor a jusante até o sensor avaliado	Utilizado em alguns modelos, com ados em Painel
Velocidade mediana no trecho a montante	Vel_TM	Velocidade mediana no trecho a montante do sensor avaliado	Utilizado em alguns modelos, com ados em Painel
Velocidade mediana no trecho a jusante	Vel_TJ	Velocidade mediana no trecho a jusante do sensor avaliado	Utilizado em alguns modelos, com ados em Painel
Inverso da Distância do sensor a montante	I-Dist_M	Inverso da Distância do sensor a montante até o sensor avaliado	Conforme descrito no item 3.3
Inverso da Distância do sensor a jusante	I-Dist_J	Inverso da Distância do sensor a jusante até o sensor avaliado	Conforme descrito no item 3.3
Inverso do Tempo de Percurso a montante	I-Tempo_TM	Inverso do Tempo de Percurso do sensor a montante até o sensor avaliado	Conforme descrito no item 3.3
Inverso do Tempo de Percurso a jusante	I-Tempo_TJ	Inverso do Tempo de Percurso do sensor a jusante até o sensor avaliado	Conforme descrito no item 3.3
Inverso da Velocidade mediana no trecho a montante	I-Vel_TM	Inverso da Velocidade mediana no trecho a montante do sensor avaliado	Conforme descrito no item 3.3
Inverso da Velocidade mediana no trecho a jusante	I-Vel_TJ	Inverso da Velocidade mediana no trecho a jusante do sensor avaliado	Conforme descrito no item 3.3
Complemento do Inverso do Tempo de Percurso a jusante	II-Tempo_TJ	Complemento do Inverso do Tempo de Percurso do sensor a jusante até o sensor avaliado	Conforme descrito no item 3.3
Complemento do Inverso da Velocidade mediana no trecho a montante	II-Vel_TM	Complemento do Inverso da Velocidade mediana no trecho a montante do sensor avaliado	Conforme descrito no item 3.3

Fonte: o autor

Importante notar que o tempo e a velocidade de percurso dependem da identificação de ao menos um veículo nos sensores de início e final do trecho. Conforme apresentado no item 4.2.4, existe limitação na qualidade de identificação de veículos, em especial no período noturno. Associando-se a isso a redução do fluxo que ocorre nas madrugadas, verificamos que não foi possível determinar valores de tempo e velocidade de percurso para cerca de 26% dos períodos.

Como no caso dos dados pontuais, foram criados trechos fictícios para possibilitar o tratamento de vizinhança dos trechos localizados nas bordas da rede. Os valores foram obtidos como sendo a média do trecho de borda e o seu primeiro vizinho. Também foram estabelecidas séries defasadas em até 8 intervalos para os dados de trecho, possibilitando a verificação do impacto de medições passadas do trecho avaliado sobre si e dos trechos a montante e jusante sobre ele.

4.4.3 Consolidação dos dados

Nesta etapa, visando os modelos baseados em Dados em Painel, foram acrescentadas 4 variáveis multiplicativas, que consideraram características pontuais e de trecho: velocidade mediana ou volume (fluxo) nos sensores vizinhos e a distância até eles (Tabela 4-6).

Tabela 4-6 – Variáveis que consideram dados pontuais e de trecho

Variável	Nome da Variável no BD	Descrição	Observações
Velocidade pela Distância do sensor a montante	Vel_RM_Dist	Produto da velocidade mediana no sensor a montante pelo inverso da Distância entre o sensor a montante e o sensor avaliado	Variável multiplicativa que associa a velocidade mediana num sensor a montante e a extensão do trecho a montante do sensor em análise.
Velocidade pela Distância do sensor a jusante	Vel_RJ_Dist	Produto da velocidade mediana no sensor a jusante pelo inverso da Distância entre o sensor a jusante e o sensor avaliado	Variável multiplicativa que associa a velocidade mediana num sensor a jusante e a distância dele até o sensor em análise
Volume pela Distância do sensor a montante	Vol_RM_Dist	Produto do volume no sensor a montante pelo inverso da Distância entre o sensor a montante e sensor o avaliado	Variável multiplicativa que associa o volume em um sensor a montante e a distância dele até o sensor em análise
Volume pela Distância do sensor a jusante	Vol_RJ_Dist	Produto do volume no sensor a jusante pelo inverso da Distância entre o sensor a jusante e sensor o avaliado	Variável multiplicativa que associa o volume em um sensor a jusante e a distância dele até o sensor em análise

Fonte: o autor

Após a conclusão da preparação dos dados pontuais (

Tabela 4-4) e de trecho (Tabela 4-5) e a criação das variáveis mistas (Tabela 4-6), todos os dados foram reunidos num único banco de dados cujas características estão apresentadas na Figura 4-7.

Figura 4-7 – Detalhe do banco de dados utilizado

Eixo	Num_Agrupado	Periodo	Est_Meteorol	Data	dia_sem	Ocup_media	Ocup_mediana
23deMaio_C:55309	10433 : 11160	P1:33231	CGE:76755	Length:198853	0:25848	Min. : 83	Min. : 83
Brasil_N :44583	10632 : 11160	P2:16614	PI :55684	Class :character	1:25919	1st Qu.: 690	1st Qu.: 677
Reboucas_B:65609	10748 : 11160	P3:33239	SA :10680	Mode :character	2:25694	Median : 795	Median : 752
Sumare_N :33352	10482 : 11159	P4:49720	VM :55734		3:32008	Mean : 1258	Mean : 840
	10554 : 11159	P5:41231			4:31918	3rd Qu.: 1268	3rd Qu.: 906
	10637 : 11159	P6:24818			5:31661	Max. :44884	Max. :53399
	(Other):131896				6:25805	NA's :780	NA's :780
Ocup_std	Volume	Vol_eq	CH_1h	data3	radar_montante	radar_jusante	
Min. : 0.0	Min. : 1	Min. : 0.33	Min. : 0.0	Min. :15198621	10528 : 22297	10433.5 : 11160	
1st Qu.: 198.0	1st Qu.: 55	1st Qu.: 54.65	1st Qu.: 0.0	1st Qu.:15205252	10509 : 21093	10622 : 11160	
Median : 324.8	Median :111	Median :111.00	Median : 0.0	Median :15211900	10482 : 11160	10748.5 : 11160	
Mean : 1673.8	Mean :136	Mean :135.09	Mean : 391.4	Mean :15211904	10667 : 11160	10433 : 11159	
3rd Qu.: 1070.7	3rd Qu.:169	3rd Qu.:167.63	3rd Qu.: 0.0	3rd Qu.:15218529	10679 : 11160	10637 : 11159	
Max. :44431.2	Max. :654	Max. :648.92	Max. :48400.0	Max. :15225403	10500 : 11159	10667 : 11159	
NA's :1293	NA's :780	NA's :780			(Other):110824	(Other):131896	
trecho_montante	trecho_jusante	Vol_RM	ocup_md_RM	Vol_RJ	Ocup_md_RJ	Tempo_TM	
3 : 11160	3.5 : 11160	Min. : 0.0	Min. : 14.0	Min. : 0.0	Min. : 49.0	Min. : 32.0	
10 : 11160	11 : 11160	1st Qu.: 53.0	1st Qu.: 675.0	1st Qu.: 53.0	1st Qu.: 542.0	1st Qu.: 114.0	
18 : 11160	18.5 : 11160	Median : 111.0	Median : 760.0	Median : 112.0	Median : 737.0	Median : 171.0	
2 : 11159	3 : 11159	Mean : 139.7	Mean : 885.4	Mean : 154.4	Mean : 899.4	Mean : 253.6	
7 : 11159	8 : 11159	3rd Qu.: 173.0	3rd Qu.: 962.0	3rd Qu.: 189.0	3rd Qu.: 924.0	3rd Qu.: 268.0	
8 : 11159	9 : 11159	Max. :8702.0	Max. :117640.0	Max. :47524.0	Max. :2548215.0	Max. :5558.0	
(Other):131896	(Other):131896	NA's :3921	NA's :3921	NA's :2595	NA's :2595	NA's :20387	
Tempo_TJ	log_ocup_mediana	log_ocup_media	Vel_Reg	Vel_media	Vel_mediana	Vel_std	
Min. : 15.0	Min. : 4.419	Min. : 4.419	Min. :50.04	Min. : 0.00	Min. : 0.00	Min. : 0.000	
1st Qu.: 109.0	1st Qu.: 6.518	1st Qu.: 6.537	1st Qu.:50.04	1st Qu.: 28.00	1st Qu.:29.00	1st Qu.: 5.522	
Median : 173.0	Median : 6.623	Median : 6.679	Median :50.04	Median : 37.00	Median :38.00	Median : 7.032	
Mean : 269.9	Mean : 6.630	Mean : 6.867	Mean :52.28	Mean : 35.16	Mean :35.26	Mean : 7.493	
3rd Qu.: 276.0	3rd Qu.: 6.810	3rd Qu.: 7.146	3rd Qu.:50.04	3rd Qu.: 41.00	3rd Qu.:41.00	3rd Qu.: 8.903	
Max. :11024.0	Max. :10.886	Max. :10.712	Max. :60.12	Max. :111.00	Max. :95.00	Max. :133.681	
NA's :34969	NA's :780	NA's :780	NA's :34969	NA's :780	NA's :780	NA's :1293	
Vel_md_RM	Vel_md_RJ	Vel_TM	Vel_TJ	Vel_reg_50	Vel_tp	dia_u	Vel_vol
Min. : 0.0	Min. : 0.00	Min. : 0.000	Min. : 0.00	Min. :25.02	0: 35469	0:141387	Min. : 0
1st Qu.: 28.0	1st Qu.: 28.00	1st Qu.: 1.000	1st Qu.: 1.00	1st Qu.:25.02	1:163384	1: 57466	1st Qu.: 1866
Median : 37.0	Median : 38.00	Median : 2.000	Median : 2.00	Median :25.02			Median : 3510
Mean : 35.3	Mean : 36.43	Mean : 2.508	Mean : 2.48	Mean :26.14			Mean : 4893
3rd Qu.: 41.0	3rd Qu.: 43.00	3rd Qu.: 4.000	3rd Qu.: 3.00	3rd Qu.:25.02			3rd Qu.: 5559
Max. :1874.0	Max. :1518.00	Max. :13.000	Max. :17.00	Max. :30.06			Max. :29541
NA's :3921	NA's :2596	NA's :20387	NA's :34969				NA's :780
Vel_oclog	oclog_vol	ocupancia	Dist_M	Dist_J	Num_faixa	Fx_Onibus	Classe
Min. : 0.0	Min. : 4.0	Min. : 85	Min. :0.601	Min. :0.601	3:100209	N: 55766	Arterial:154701
1st Qu.:209.0	1st Qu.: 370.0	1st Qu.: 41515	1st Qu.:1.170	1st Qu.:1.149	4: 65166	S:143087	Expressa: 44152
Median :256.0	Median : 780.0	Median : 100163	Median :1.367	Median :1.272	5: 22319		
Mean :240.6	Mean : 919.6	Mean : 153511	Mean :1.497	Mean :1.454	6: 11159		
3rd Qu.:274.0	3rd Qu.:1203.0	3rd Qu.: 226295	3rd Qu.:2.005	3rd Qu.:1.733			
Max. :457.0	Max. :4051.0	Max. :1127451	Max. :2.756	Max. :2.756			
NA's :780	NA's :780	NA's :780					
data4	data5	I_Vol_RM	I_Vol_RJ	I_Tempo_TM	I_Tempo_TJ	II_Tempo_TJ	
15198621: 18	(15198621): 18	Min. :0.000	Min. :0.0000	Min. :0.000	Min. :0.0000	Min. : -Inf	
15198624: 18	(15198624): 18	1st Qu.:0.006	1st Qu.:0.0053	1st Qu.:0.006	1st Qu.:0.0053	1st Qu.:0.9811	
15198626: 18	(15198626): 18	Median :0.009	Median :0.0089	Median :0.009	Median :0.0089	Median :0.9911	
15198628: 18	(15198628): 18	Mean : Inf	Mean : Inf	Mean : Inf	Mean : Inf	Mean : -Inf	
15198631: 18	(15198631): 18	3rd Qu.:0.019	3rd Qu.:0.0189	3rd Qu.:0.019	3rd Qu.:0.0189	3rd Qu.:0.9947	
15198633: 18	(15198633): 18	Max. : Inf	Max. : Inf	Max. : Inf	Max. : Inf	Max. :1.0000	
(Other):198745	(Other):198745	NA's :3921	NA's :2595	NA's :3921	NA's :2595	NA's :2595	
I_Vel_TM	I_Vel_TJ	II_Vel_TM	I_Dist_M	I_Dist_J	Vel_RM_Dist	Vel_RJ_Dist	
Min. :0.077	Min. :0.06	Min. : -Inf	Min. :0.03628	Min. :0.03628	Min. : 0.000	Min. : 0.000	
1st Qu.:0.250	1st Qu.:0.33	1st Qu.:0.000	1st Qu.:0.04988	1st Qu.:0.05770	1st Qu.: 1.794	1st Qu.: 1.833	
Median :0.500	Median :0.50	Median :0.500	Median :0.07315	Median :0.07862	Median : 2.366	Median : 2.516	
Mean : Inf	Mean : Inf	Mean : -Inf	Mean :0.08062	Mean :0.08230	Mean : 2.758	Mean : 2.931	
3rd Qu.:1.000	3rd Qu.:1.00	3rd Qu.:0.750	3rd Qu.:0.08547	3rd Qu.:0.08703	3rd Qu.: 3.438	3rd Qu.: 3.610	
Max. : Inf	Max. : Inf	Max. :0.923	Max. :0.16639	Max. :0.16639	Max. :85.027	Max. :130.189	
NA's :20387	NA's :34969	NA's :20387			NA's :3921	NA's :2596	
Vol_RM_Dist	Vol_RJ_Dist						
Min. : 0.000	Min. : 0.000						
1st Qu.: 3.610	1st Qu.: 4.088						
Median : 7.818	Median : 9.622						
Mean : 10.112	Mean : 10.475						
3rd Qu.: 14.673	3rd Qu.: 14.739						
Max. :394.828	Max. :3736.164						
NA's :3921	NA's :2595						

Fonte: o autor

Verifica-se que o banco de dados estava formado com 198.853 linhas de dados, cada qual com 62 colunas. As variáveis categóricas são indicadas com suas categorias e quantidades e as quantitativas são indicadas de acordo com os quartis, a mediana e a quantidade de registros sem dados (NA). A posterior inclusão das variáveis defasadas em 8 intervalos acresceu 294 colunas ao banco de dados.

Nos estudos com Dados em Painel não podem ser utilizados registros onde a variável dependente não contenha dados. Assim, o último passo antes da divisão do banco de dados entre treino e teste foi a eliminação de 780 registros sem a informação de velocidade no sensor. Dessa forma o banco de dados final estava formado por 198.073 linhas de dados.

A divisão em treino e teste resultou em bancos de dados com, respectivamente, 122.524 e 75.549 registros. No entanto, os dados defasados acabam por gerar a perda dos 8 primeiros registros da série de cada local estudado. Assim, a quantidade final utilizada no treino foi de 121.245 registros. Para o teste foram utilizados 74.663 registros.

Por fim, mesmo com a complementação dos dados de borda, os modelos de Dados em Painel, que envolveram sensores vizinhos, foram elaborados com 118.740 registros, ou seja uma redução próxima a 2%. Para os modelos que trataram com os trechos vizinhos essa redução foi da ordem de 26%, de modo que foram utilizados 90.161 registros. Essa redução maior para o caso dos trechos decorre da grande quantidade de valores de tempo e velocidade de percurso ausentes.

4.5 ANÁLISE DE DADOS

O processo de Análise de Dados faz parte da metodologia, uma vez que parte das informações recebidas podem ser valores atípicos (*“outliers”*), o que poderia enviesar as estimações. Por outro lado, fatores externos - como condições meteorológicas e incidentes - que ocorrem no sistema viário, impactam a fluidez e capacidade do tráfego, podendo impactar os valores das características do tráfego.

A Análise dos dados ocorre em dois momentos:

- a) Análise dos dados na etapa de pré-processamento, verificando se dentre os dados existe um dado extremo, que pode enviesar o valor determinado para o intervalo, num determinado sensor. Caso haja, o mesmo pode ser excluído, prosseguindo-se com o restante dos dados. Isto já foi feito de modo limitado na limpeza inicial dos

dados de estudo, quando foram excluídos 4 valores de velocidades acima de 90 km/h (1,5 vezes a máxima permitida), que constavam nos dados originais;

- b) Verificação de valores atípicos nas séries temporais, o que pode ser causado por falhas no equipamento e incidentes na via, dentre outros. Neste caso o tratamento pode ser o de substituir o valor medido pelo valor estimado no intervalo anterior.

Pela simplicidade de análise e aplicação direta aos dados, para a identificação de valores atípicos, foram considerados o teste GESD e o do Quartil, aplicados tanto nos valores recebidos dos sensores como nos erros obtidos nas estimações.

4.5.1 Teste GESD

O teste de Grubbs pode ser descrito pela Expressão 4-2. Foi determinado para analisar se o maior/menor valor de uma amostra é ou não atípico. Segundo Grubbs (1969), quando o desvio padrão é bem conhecido é possível utilizar seu método, sem a necessidade de novo cálculo do desvio padrão a cada interação. Nesta tese é possível determinar o desvio padrão de cada local com base no banco de dados histórico disponível, sendo que o cálculo pode ser feito periodicamente para atualização do valor.

$$G = \frac{|\bar{O} - P_i|}{S} \quad (4-2)$$

Onde: G = Estatística do teste de Grubbs, \bar{O} = valor médio da variável, P_i = valor avaliado, S = desvio padrão. O resultado é comparado com o valor crítico de Grubbs, disponível em (WILRICH, 2011).

No método GESD (ROSNER, 1983), que consiste numa generalização da aplicação do método de Grubbs, a mesma equação é utilizada, todavia o resultado é comparado com a “distribuição T”. Esta generalização se apoia em evitar um grande número de interações utilizando-se o desvio padrão constante.

4.5.2 Teste do quartil

Na abordagem pelo método do Quartil, segundo Kannan (2013), calculados os quartis Qt_1 (25%), Qt_2 (50%) e Qt_3 (75%), pode-se determinar um valor atípico pela Expressão 4-3 e valores extremos com base na Expressão 4-4, sendo que nos dois casos $h = Qt_3 - Qt_1$:

$$x_a < Qt_1 - 1.5 \cdot h \quad \text{ou} \quad x_a > Qt_3 + 1.5 \cdot h \quad \text{valor atípico} \quad (4-3)$$

$$x_e < Qt_1 - 3 \cdot h \quad \text{ou} \quad x_e > Qt_3 + 3 \cdot h \quad \text{valor extremo} \quad (4-4)$$

4.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO 4

Neste capítulo foram apresentados os dados utilizados na tese e os processos de tratamento inicial, que possibilitam a obtenção de séries temporais organizadas, com base nos sensores e nos trechos.

Foram apresentadas algumas das características dos dados, no tocante à composição do fluxo, como concentração da presença de automóveis, flutuação sazonal ao longo das horas do dia e situações atípicas.

Os dados provenientes dos equipamentos de fiscalização instalados no Município de São Paulo estavam fechados e inacessíveis à pesquisadores e desenvolvedores. O tratamento efetuado nesta pesquisa permitiu a anonimização da informação sensível, tornando esses dados acessíveis para o desenvolvimento de pesquisas e aplicativos de gestão do tráfego.

A dificuldade enfrentada com relação à identificação dos veículos foi apresentada, todavia as análises mostraram a possibilidade de obtenção dos tempos de percurso para o modelo, havendo alguma restrição quanto a essa definição para o período da madrugada. Esse fato não é crítico, já que na madrugada o tráfego é bastante reduzido e os tempos de percurso se aproximam da situação de fluxo livre, sendo possível adotar valores relativos à velocidade regulamentada na via.

Os dados de chuva, compostos pelo monitoramento dos índices pluviométricos foram correlacionados com os sensores, possibilitando sua aplicação em modelos multivariados.

O processo de montagem das séries temporais pontuais e de trecho foi detalhado. Destaca-se que esse detalhamento foi utilizado, com sucesso, como base na montagem dos scripts para sua execução.

Por fim, foi apresentado o método de análise de dados que ocorre em dois momentos da metodologia proposta. O método foi elaborado de modo a permitir uma análise simples e rápida dos dados, utilizando testes já utilizados e apresentados pela literatura.

5 APLICAÇÃO E RESULTADOS

Neste capítulo são apresentados os resultados dos estudos efetuados nesta tese. Os estudos estão embasados na revisão bibliográfica apresentada no capítulo 2, na metodologia proposta e discutida no capítulo 3 e no tratamento de dados, apresentado no capítulo 4.

Os estudos foram divididos em 4 etapas principais. Numa etapa preliminar foi feita a abertura e análise inicial dos dados dos sensores (radares), já descrita no capítulo 4.

Na etapa seguinte foi efetuado um estudo sobre a possibilidade de utilização de STARIMA na estimação do tráfego, onde se detectou a necessidade de estudos mais detalhados com relação às matrizes espaciais presentes no modelo.

Num terceiro momento, foram efetuados os estudos de previsão com aplicação de Dados em Painel com estruturas longitudinal e aninhada, tendo sido elaborados modelos de previsão da velocidade pontual.

Por fim, foram efetuadas comparações dos resultados obtidos com os dois métodos.

Em paralelo aos estudos, foram elaborados modelos utilizando Dados em Painel para a predição de fluxo (volume) numa rede similar à utilizada nesta tese. Este estudo, feito em conjunto com a equipe de planejamento da CET, teve por objetivo preencher lacunas nos dados horários de fluxo obtidos com base nos radares de fiscalização. Este estudo complementar está apresentado, de modo resumido, no Apêndice B.

5.1 APLICAÇÃO

Esta Seção está dividida em 3 partes, sendo iniciada pela apresentação dos estudos sobre matrizes de relação espacial, que precederam os trabalhos de estimação dos modelos baseados em STARIMA, apresentada na segunda parte. Na sequência, apresentamos a aplicação dos modelos de Regressões com Dados em Painel, Longitudinal e HLM.

5.1.1 Estimação das características do tráfego com STARIMA

A aplicação do modelo STARIMA foi precedida da criação de uma ferramenta específica para sua aplicação. O desenvolvimento da ferramenta se deu através de um grupo de estudantes de graduação da Poli/USP, que desenvolveu o aplicativo durante os estudos de seu Trabalho de Final de Curso (PAZ; TRINTINALIA, 2021).

O estudo foi dividido em dois momentos, o primeiro no qual foi elaborado um estudo específico sobre as características das matrizes de relação espacial, utilizando-se os dados de volume de tráfego, conforme apresentado a seguir.

Num segundo momento, o conhecimento adquirido foi aplicado nos estudos relativos à velocidade pontual, utilizados na comparação com os modelos de Dados em Painel. As matrizes escolhidas para esta etapa foram as que consideram a distância entre sensores e a que considera o tempo de percurso. A primeira matriz foi a escolhida por não necessitar da elaboração de séries temporais para os trechos, sendo assim mais simples de aplicar. A segunda por ser uma das que apresentaram melhor desempenho nos estudos prévios sobre a aplicação de matrizes.

Destaca-se que tanto nos modelos relativos à velocidade como nos que estimaram o volume, as mesmas matrizes de relação espacial foram utilizadas, pois a forma de aplicação do modelo permite a determinação de diversas características com uma mesma base de matrizes. Apenas é necessária a criação das séries temporais separadas e específicas de cada característica de tráfego estudada, seja esta a velocidade, o fluxo (volume), ou a ocupação.

A aplicação foi feita com base no aplicativo STARIMA (PAZ; TRINTINALIA, 2021), desenvolvido em R/Shiny. A aplicação foi feita diretamente na nuvem, no site da shinyapps.io, em computador virtual com 8 Gb de RAM.

O processamento não chegou a ser cronometrado, porém, verificamos que a etapa de montagem e calibração foi a que mais consumiu tempo e esforço computacional, sendo dependente do tamanho da rede analisada. Assim, nos modelos com todos os 18 sensores, o tempo de montagem do modelo chegava a superar 30 min. Todavia, pela pequena quantidade de parâmetros gerados, a estimação com a base de testes se mostrou mais ágil, indicando a possibilidade de sua utilização em estimações em tempo próximo ao real.

5.1.1.1 Análise das matrizes de relação espacial – variável volume

Para os estudos de matrizes foram elaborados modelos comparativos, tendo como base a característica volume de tráfego (fluxo) cuja aplicação - na rede de estudo - propiciou melhor entendimento sobre as características a serem utilizadas. As comparações envolveram 5 aspectos distintos e elencados a seguir:

- a. Pesos = 8 tipos: sendo 5 simétricos UN, 1/D, 1/T, 1/V e 1/L e 3 assimétricos 1-1/T, 1-1/V, e 1-1/L. conforme descritos no item 3.3.1. Visando verificar se os modelos assimétricos, menos presentes na literatura, propiciariam maior acurácia;
- b. Vizinhança = 3 tipos: agrupada numa única ordem (AG); várias ordens com vizinhança por contiguidade (VC); várias ordens com vizinhança pelo Lag de Percurso (LP). Conforme descritas no item 3.3.2, com o objetivo de verificar a possibilidade de uso do modelo AG que é mais simples;
- c. Abrangência = 3 grupos: Análise de 1 eixo expresso com 4 sensores (23 de Maio) e 1 eixo arterial com 6 sensores (Rebouças/Consolação) e da rede com 18 sensores (inclui os 2 eixos examinados isoladamente). Detalhados no item 4.1, a fim de verificar o comportamento do modelo, ao reunir numa única rede, eixos de tráfego com características diferentes;
- d. Horizonte de Previsão = 3 horizontes: 1, 5 e 15 passos à frente, ou seja, 4, 20 e 60 minutos respectivamente, numa análise da possibilidade de uso do modelo STARIMA para os processos de previsão de tráfego;
- e. Característica do Tráfego no Período = 3 Características: Típico (fluxo semelhante ao de treino), Atípico A (sob impacto de chuvas fortes com forte redução no fluxo) e Atípico B (véspera de feriado, 6 meses à frente do treino). Neste caso, buscou-se verificar a acurácia do modelo, quando da ocorrência de situações diferentes das que foram utilizadas para criá-lo.

As matrizes foram construídas com base nos dados reais de campo, incluindo a distância de trajeto entre os sensores, o tempo e a velocidade média de percurso. A Figura 5-1 apresenta uma das matrizes gerais assimétricas (1-1/T), utilizada nos testes, sendo possível observar os valores relacionados com o tempo de percurso entre os diversos sensores.

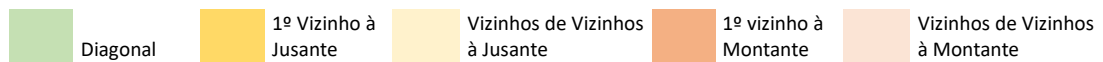
Para a montagem das matrizes foram utilizados os dados de fluxo e velocidade média do período das 12h às 16h, evitando os períodos da madrugada, que tem baixo fluxo, e dos picos,

cuja característica pode variar mais em decorrência dos congestionamentos. Os testes foram realizados para o período das 15h às 17h.

Mantendo-se a devida parcimônia, entendemos por bem não elaborar modelos onde a matriz de correlação variasse a cada passo, de modo que não foram testados modelos com matrizes dinâmicas.

Figura 5-1 – Matriz Geral de Pesos 1-1/T utilizada nos testes

	10426	10484	10492	10521	10507	10516	10528	10554	10637	10667	10632	10622	10563	10509	10556	10591	10594	10679	10748
10426	1	0,837	0,917	0,932	0,850	0	0,890	0,908	0,953	0,961	0,964	0,966	0,924	0	0	0	0	0	0
10484	0,163	1	0,833	0,884	0	0	0,663	0,788	0,933	0,948	0,954	0,958	0,856	0	0	0	0	0	0
10492	0,083	0,167	1	0,622	0	0,303	0	0	0	0	0	0	0	0	0	0	0	0	0
10521	0,068	0,116	0,378	1	0	0,168	0	0	0	0	0	0	0	0	0	0	0	0	0
10507	0,150	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10516	0	0	0,697	0,832	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
10528	0,110	0,337	0	0	0	0	1	0,430	0,916	0,939	0,947	0,952	0,749	0	0	0	0	0	0
10554	0,092	0,212	0	0	0	0	0,570	1	0,902	0,932	0,941	0,947	0	0	0	0	0	0	0
10637	0,047	0,067	0	0	0	0	0,084	0,098	1	0,775	0,852	0,885	0	0	0	0	0	0	0
10667	0,039	0,052	0	0	0	0	0,061	0,068	0,225	1	0,571	0,766	0	0	0	0	0	0	0
10632	0,036	0,046	0	0	0	0	0,053	0,059	0,148	0,429	1	0,484	0	0	0	0	0	0	0
10622	0,034	0,042	0	0	0	0	0,048	0,053	0,115	0,234	0,516	1	0	0	0	0	0	0	0
10563	0,076	0,144	0	0	0	0	0,251	0	0	0	0	0	1	0	0	0	0	0	0
10509	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0,777	0,847	0,878	0,949	0,968
10556	0	0	0	0	0	0	0	0	0	0	0	0	0	0,223	1	0,511	0,733	0,934	0,962
10591	0	0	0	0	0	0	0	0	0	0	0	0	0	0,153	0,489	1	0	0	0
10594	0	0	0	0	0	0	0	0	0	0	0	0	0	0,122	0,267	0	1	0,912	0,956
10679	0	0	0	0	0	0	0	0	0	0	0	0	0	0,051	0,066	0	0,088	1	0,912
10748	0	0	0	0	0	0	0	0	0	0	0	0	0	0,032	0,038	0	0,044	0,088	1



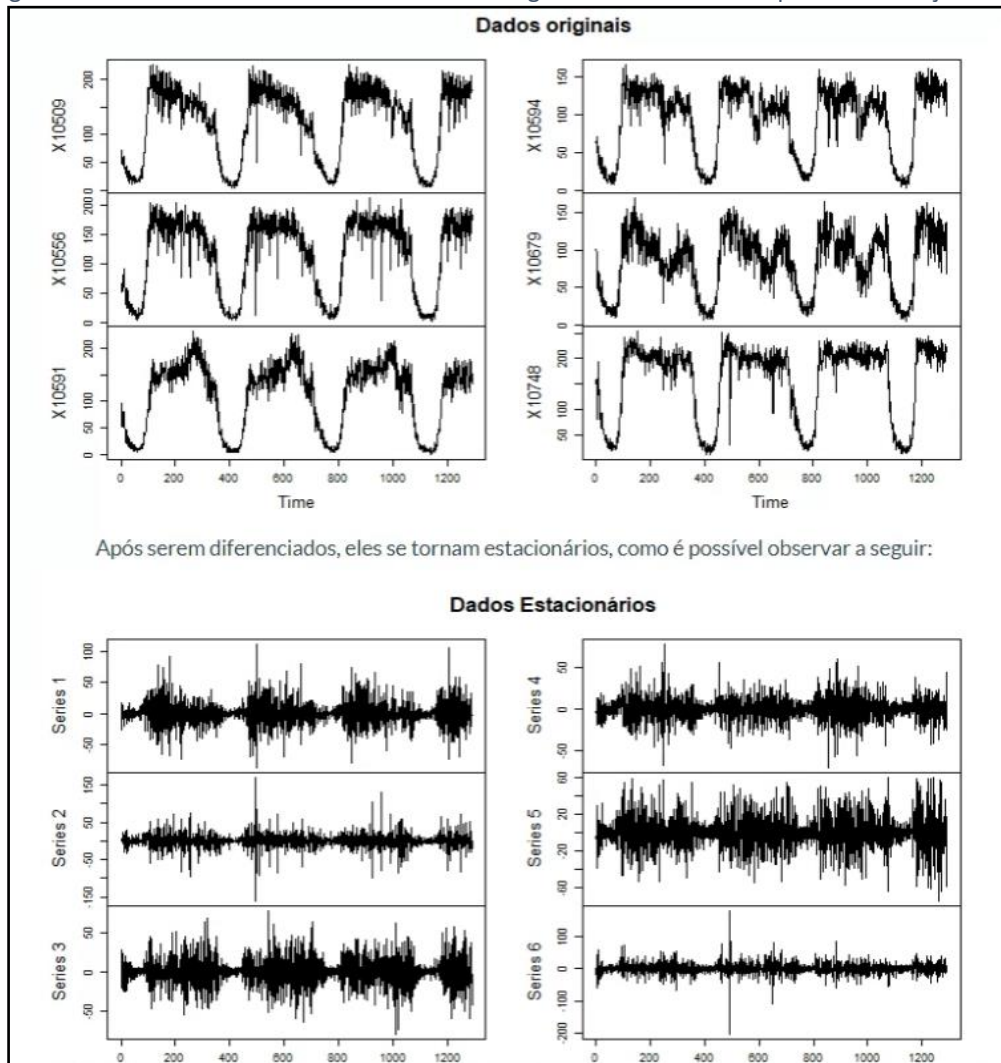
Fonte: o autor

Os testes das matrizes e a aplicação do modelo STARIMA foram realizados com a utilização do aplicativo desenvolvido em R Shiny, dedicado à análise do uso do modelo STARIMA em corredores e redes de tráfego (PAZ; TRINTINALIA, 2021), apresentado no Apêndice D. Salienta-se que o aplicativo foi desenvolvido durante um trabalho de conclusão de curso (TCC) sob orientação do autor desta pesquisa.

A construção do modelo é feita com base em (PFEIFER; DEUTSCH, 1980a), sendo que o aplicativo citado define o valor da ordem de integração d necessária para garantir a estacionariedade das séries temporais - Figura 5-2. Na mesma imagem verifica-se que a dispersão dos valores apresenta diferenças, num mesmo corredor. As séries 2 e 6 apresentam menor dispersão dos dados estacionários, nas demais a dispersão é maior ao longo do dia e menor durante a madrugada. Embora este fato não impeça a aplicação do modelo STARIMA,

pode ocorrer impacto negativo na qualidade da previsão. Numa situação em que as flutuações fossem mais uniformes, a influência de um local em outro seria mais homogênea, favorecendo a aplicação do modelo STARIMA.

Figura 5-2 – Característica dos dados de treino originais e estacionários após diferenciação



Fonte: Extraído do aplicativo STARIMA (PAZ; TRINTINALIA, 2021)

A escolha inicial das ordens da autocorrelação q e m é feita a partir de inspeção visual da aplicação da função Space Time Autocorrelation Function (STACF) e as ordens de p e λ , da média móvel, baseiam-se na função Space Time Partial Autocorrelation Function (STPACF), já apresentadas no item 3.2.1.

A definição final das ordens foi feita através de um procedimento de análise exaustiva de todos os modelos possíveis, tendo como limite máximo as ordens de $p+1$, $\lambda+1$, $q+1$ e $m+1$, sendo p , λ , q e m as ordens obtidas nos gráficos (Tabela 5-1). As ordens espaciais foram limitadas a 1 para as matrizes agrupadas, uma vez que nos cenários deste tipo apenas a matriz

de ordem 1 é utilizada. Para os cenários com matrizes desagrupadas, com base na parcimônia recomendada por Pfeifer e Deutsch (1980a), foram utilizadas matrizes até à 4ª ordem espacial. E, também de modo parcimonioso, para todos os modelos as ordens temporais foram limitadas a 4. Por fim, os parâmetros são calculados pelo aplicativo, a partir da aplicação do método de Levenberg Marquadt.

Tabela 5-1 – Exemplo de calibração exaustiva de um modelo

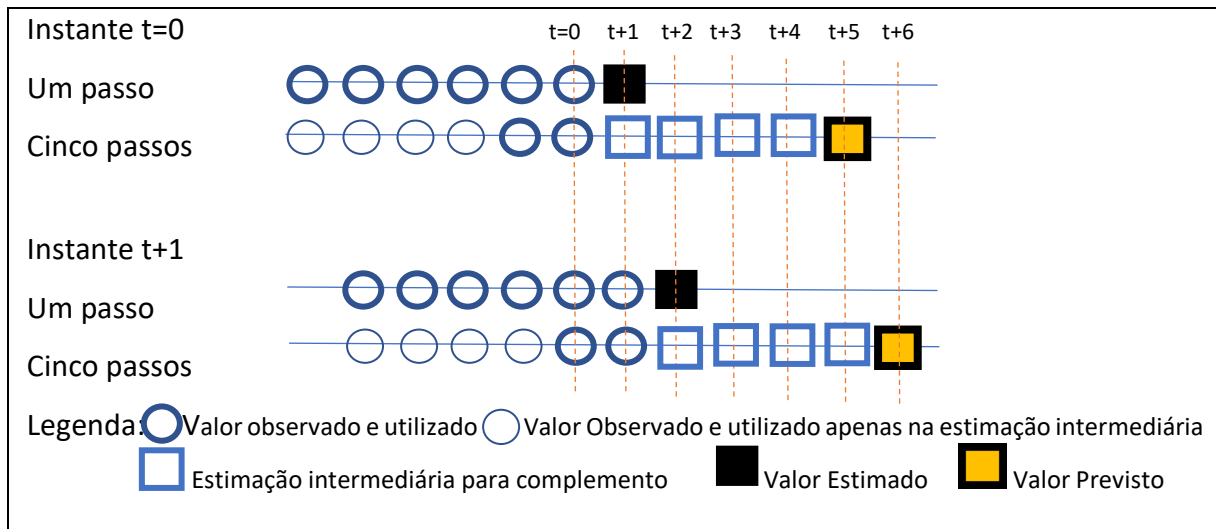
ORDENS					INDICADORES DE DESEMPENHO			
d	p	λ	q	m	E	RSS	MAE	Box Pierce
1	3	3	1	1	0.979	5.68E+05	1.53E+01	16.37
1	3	2	1	1	0.979	5.69E+05	1.53E+01	16.71
1	3	1	1	1	0.979	5.70E+05	1.53E+01	17.42
1	3	3	1	0	0.979	5.70E+05	1.53E+01	17.39
1	3	2	1	0	0.979	5.71E+05	1.53E+01	17.43
1	3	1	1	0	0.979	5.72E+05	1.53E+01	18.73

Fonte: Extraído do aplicativo STARIMA (PAZ; TRINTINALIA, 2021)

Quanto às ordens obtidas e aplicadas nos modelos, temos que a diferenciação ficou fixa em 1, sendo suficiente para tornar as séries estacionárias. Já as ordens espaciais e temporais utilizadas - p , λ , q e m - na obtenção dos parâmetros dos modelos, variaram de modelo para modelo. Para os modelos separados: p variou entre 1 e 2, o λ utilizado foi 1 (nos modelos AG) e 3 (nos demais modelos), q ficou entre 2 e 3 e m entre 0 e 1. Os modelos da rede foram estabelecidos com p e q igual a 4 e m foi 0, na maior parte dos casos, λ foi 1, sendo 2 apenas nos modelos LP.

Os testes foram realizados considerando os horizontes de 1, 5 e 15 intervalos à frente. O processo de rolagem dos dados foi aplicado, de modo que a cada passo, o dado mais antigo era descartado e um dado novo era acrescentado, conforme exemplificado na Figura 5-3

Figura 5-3 – Processo de Rolagem Iterativa de Dados para Estimação com STARIMA



Fonte: adaptado de (ZHAN et al., 2019)

Inicialmente foram testados os 2 Eixos de tráfego, para cada um dos 23 tipos de matriz de peso propostos. Na sequência foram elaboradas previsões para a Rede de 18 radares, com as 7 matrizes agrupadas. Por fim, foram testados os modelos 1-1/T e 1-1/V desagrupados (VC e LP) para toda a rede. Assim, o estudo comparou 57 previsões para cada um dos 3 horizontes, e para as 3 características de tráfego, totalizando 513 previsões.

Na análise inicial, com os 2 eixos, buscou-se comparar a acurácia dos modelos em duas vias com características diferentes de estrutura e fluxo de veículos:

- Av. 23 de Maio no sentido do centro da cidade. Via expressa com 4 faixas, que conta com 4 radares e cujos congestionamentos se concentram no horário de pico da manhã, com característica semelhante aos níveis de serviço E e F (DNIT MINISTÉRIO DOS TRANSPORTES DEPARTAMENTO NACIONAL DE INFRA-ESTRUTURA DE TRANSPORTES DIRETORIA DE PLANEJAMENTO E PESQUISA COORDENAÇÃO-GERAL DE ESTUDOS E PESQUISA INSTITUTO DE PESQUISAS RODOVIÁRIAS, 2006). No restante do período do dia o tráfego flui com certas restrições, nível de serviço variando entre C e D;
- Av. Rebouças e R. Consolação no sentido do bairro. Via arterial com 3 faixas, sendo a faixa da esquerda utilizada exclusivamente por ônibus e taxis. Conta com 6 radares, e a característica de permanecer com o fluxo carregado com alto grau de atrasos ao longo de todo o dia.

Num segundo momento, verificou-se a aplicação do modelo na rede com 18 sensores, a fim de verificar a influência da aplicação do modelo STARIMA numa rede com vias que tem perfis de tráfego diferentes, uma vez que o modelo STARIMA consiste na determinação de um modelo único para todos os pontos analisados e diferenças pontuais podem gerar inconsistências e perda de acurácia (CHENG et al., 2014).

A análise comparativa de períodos com fluxo típico e atípico é relevante, haja vista que o modelo STARIMA, ao estabelecer parâmetros únicos para toda a rede, pressupõe a uniformidade na variação dos volumes entre sensores consecutivos, o que de fato ocorre em maior quantidade no caso de dias típicos. Todavia, não considerar a possibilidade de incidentes, implica em desconsiderar o impacto gerado por eles na qualidade do fluxo em cada trecho.

Conforme verifica-se na Tabela 5-2, que apresenta um recorte dos resultados do teste RMSE relativo à previsão feita com toda a rede para o dia 06/09, com horizonte de 60 min, os cenários apresentaram acurácia relativamente semelhante para cada radar (coluna). Todavia, não se verificaram cenários totalmente ruins ou bons, havendo os que apresentaram melhor acurácia de previsão num radar e acurácia ruim noutro. Por exemplo, o tipo Rede_1/D_AG apresenta a melhor acurácia para o sensor 10591 e a pior para o sensor 10556. A variação de valores entre radares está relacionada principalmente ao volume que passa em cada local, mas também insere variação na acurácia, para cada local, decorrente da estimação com um único modelo para todos os locais.

Tabela 5-2–Recorte dos resultados de testes feitos com toda a Rede – RMSE

Matriz	Via Expressa											Via Arterial						
	10531	10500	10482	10433	10475	10528	10554	10637	10667	10632	10622	10563	10509	10556	10591	10594	10679	10748
Rede 1-1/L AG	27,9	44,2	106,0	67,0	26,4	40,1	17,8	34,5	26,6	22,1	8,3	33,0	16,3	24,4	52,7	32,9	18,7	17,7
Rede 1-1/T AG	27,4	42,8	106,7	69,4	26,1	39,1	16,8	33,6	26,0	22,8	7,6	31,0	16,2	24,7	52,9	33,1	19,0	17,9
Rede 1-1/V AG	26,7	40,5	106,1	64,2	26,8	40,7	19,3	34,8	28,9	24,7	14,0	34,6	16,3	24,5	52,4	32,3	18,2	18,2
Rede 1/D AG	27,7	43,2	105,6	67,8	27,1	46,5	24,0	38,0	31,9	24,6	13,3	34,5	17,6	32,5	51,7	32,4	18,8	18,0
Rede 1/L AG	29,4	47,6	103,9	66,6	29,7	44,0	18,5	35,4	29,7	23,1	10,6	34,3	16,2	25,1	52,8	33,2	18,8	17,4
Rede 1/T AG	26,9	41,6	102,9	67,2	28,4	38,3	17,6	33,8	27,4	21,6	7,9	33,4	16,0	25,9	53,2	34,3	18,4	17,8
Rede 1/V AG	27,4	43,5	106,6	68,7	26,3	41,8	21,4	39,0	31,6	24,5	11,1	33,3	17,8	26,2	52,5	32,2	20,5	21,0
Rede 1-1/T VC	28,5	49,3	102,1	66,4	31,0	40,4	16,9	33,5	26,6	22,6	7,6	32,6	16,9	25,7	52,9	33,3	18,6	18,0
Rede 1-1/T LP	28,3	42,5	102,1	66,2	33,0	42,9	16,4	33,3	26,5	22,4	7,6	32,4	16,0	25,5	52,6	33,2	18,8	17,9
Rede 1-1/V VC	26,6	41,5	104,8	61,9	27,1	42,1	18,1	33,8	31,5	21,8	8,1	37,7	16,9	25,2	52,5	34,1	19,1	17,4
Rede 1-1/V LP	26,8	40,7	103,8	62,7	26,5	42,1	19,5	32,8	26,6	21,2	14,7	36,6	16,2	25,1	52,2	32,9	18,9	17,6

Fonte: o autor

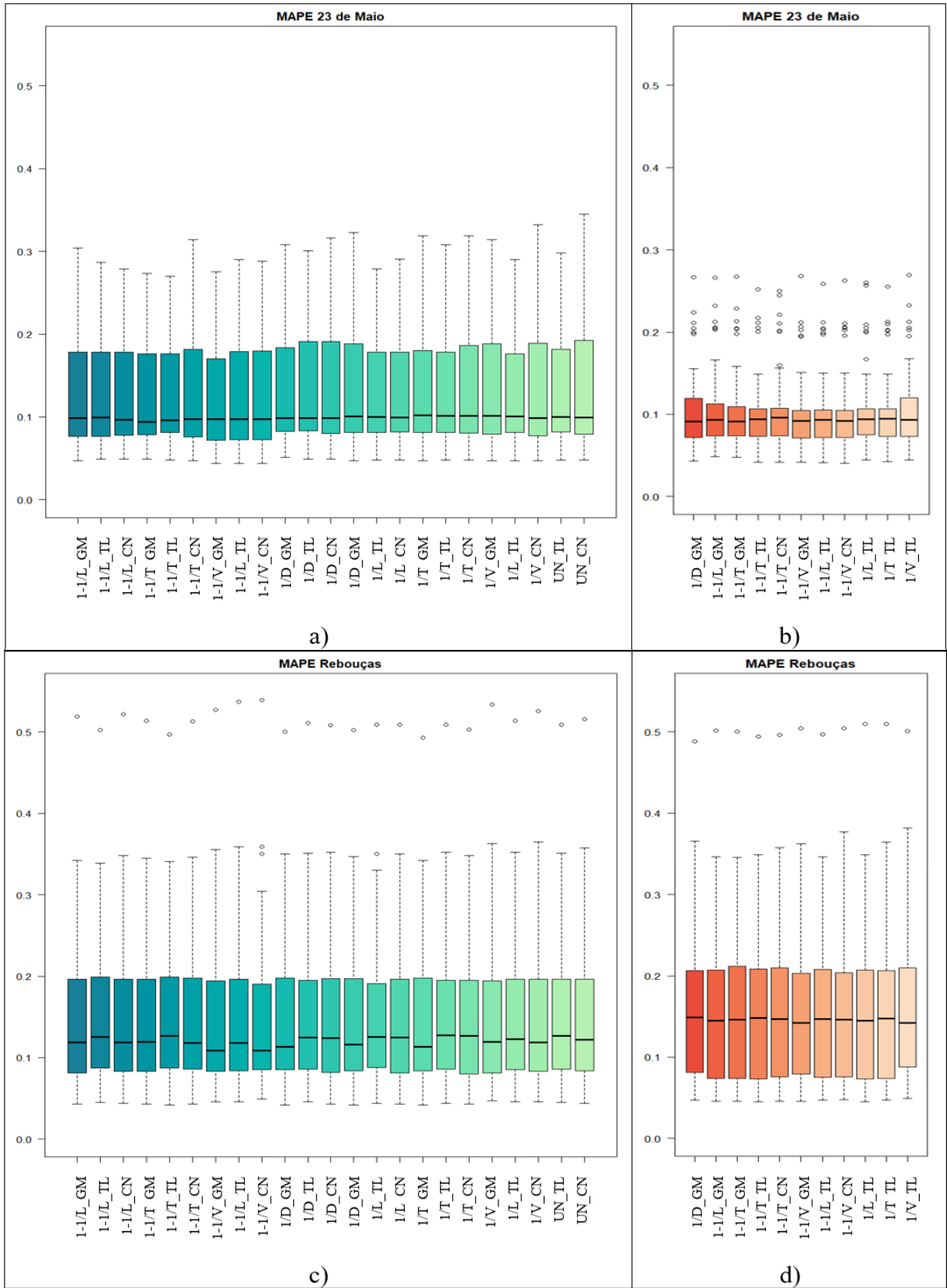
Assim, a classificação dos modelos se deu com base na média dos resultados obtidos nos dias típicos e atípicos, assim como também, tendo por base a amplitude verificada nos resultados. Foram preferidos aqueles que apresentaram melhor precisão, ou seja, com menor

desvio em relação à média, o que é possível observar nos gráficos box-plot apresentados na Figura 5-4.

Os resultados médios de acurácia dos 34 cenários testados estão apresentados na Tabela 5-3, de acordo com o MAPE, e na Tabela 5-4, de acordo com o RMSE. Os melhores resultados para cada sensor estão destacados em vermelho. Em negrito estão os resultados com a segunda melhor acurácia. Observa-se que são destacados tanto os resultados obtidos nos testes com abrangência de rede como os obtidos nas análises dos eixos de tráfego em separado. Apenas os radares dos dois eixos (Rebouças/Arterial e 23 de Maio/Expressa) foram considerados nestas comparações.

Os resultados apresentados na Tabela 5-3 e na Tabela 5-4, são discutidos ao longo deste capítulo.

Figura 5-4 – Acurácia de Previsão dos pesos testados nos corredores em separado e com toda a rede - STARIMA



a) Via Expressa em separado; b) Via Expressa no modelo com toda a rede; c) Via Arterial em separado; d) Via Arterial no modelo com toda a rede.

Fonte: o autor

Tabela 5-3 – Análise de Acurácia dos Modelos Testados – MAPE

MAPE		Típico						Atípico A						Atípico B					
Teste	Tipo	Arterial			Expressa			Arterial			Expressa			Arterial			Expressa		
		LAG 1	LAG 5	LAG 15	LAG 1	LAG 5	LAG 15	LAG 1	LAG 5	LAG 15	LAG 1	LAG 5	LAG 15	LAG 1	LAG 5	LAG 15	LAG 1	LAG 5	LAG 15
Teste por Eixo	UN_ET	0,1084	0,1201	0,1541	0,0739	0,0985	0,1083	0,1333	0,2212	0,3327	0,7755	3,6819	4,8379	0,1423	0,1646	0,2258	0,1467	0,1485	0,2021
	UN_VC	0,1078	0,1190	0,1517	0,0723	0,1043	0,1112	0,1294	0,2229	0,3343	0,9731	3,5236	4,1414	0,1431	0,1662	0,2254	0,1478	0,1536	0,2091
	1/D_AG	0,1096	0,1206	0,1526	0,0716	0,1021	0,1074	0,1316	0,2203	0,3369	1,0510	3,7667	4,8537	0,1429	0,1653	0,2238	0,1488	0,1545	0,2069
	1/D_ET	0,1087	0,1199	0,1541	0,0732	0,1004	0,1129	0,1337	0,2208	0,3333	0,9819	3,8349	5,0755	0,1423	0,1660	0,2276	0,1495	0,1507	0,2071
	1/D_VC	0,1081	0,1202	0,1526	0,0718	0,1017	0,1100	0,1299	0,2240	0,3349	1,0416	3,7807	4,8629	0,1437	0,1645	0,2237	0,1514	0,1533	0,2090
	1/L_AG	0,1095	0,1210	0,1532	0,0721	0,0999	0,1089	0,1318	0,2198	0,3377	0,8631	3,5669	4,4414	0,1429	0,1644	0,2242	0,1496	0,1526	0,2081
	1/L_ET	0,1087	0,1202	0,1533	0,0738	0,0974	0,1056	0,1330	0,2202	0,3332	0,8588	3,6905	4,7292	0,1436	0,1652	0,2262	0,1466	0,1464	0,1931
	1/L_VC	0,1081	0,1205	0,1528	0,0730	0,1014	0,1039	0,1300	0,2242	0,3354	0,9068	3,8001	5,2727	0,1436	0,1639	0,2240	0,1470	0,1521	0,2006
	1/T_AG	0,1104	0,1216	0,1545	0,0713	0,1005	0,1070	0,1334	0,2198	0,3421	0,8379	3,5704	4,5123	0,1432	0,1646	0,2250	0,1467	0,1532	0,2068
	1/T_ET	0,1087	0,1199	0,1543	0,0729	0,0971	0,1094	0,1339	0,2211	0,3339	0,7637	3,2819	3,8793	0,1420	0,1651	0,2263	0,1438	0,1490	0,2012
	1/T_VC	0,1083	0,1212	0,1529	0,0731	0,1010	0,1084	0,1305	0,2250	0,3365	0,8201	3,5885	4,5749	0,1440	0,1634	0,2227	0,1476	0,1519	0,2065
	1/V_AG	0,1068	0,1182	0,1510	0,0719	0,0978	0,1132	0,1282	0,2210	0,3331	1,0751	3,3014	3,5135	0,1436	0,1700	0,2297	0,1494	0,1475	0,2010
	1/V_ET	0,1084	0,1199	0,1537	0,0733	0,0984	0,1072	0,1328	0,2205	0,3322	0,8019	3,6513	4,7018	0,1426	0,1662	0,2280	0,1439	0,1472	0,1987
	1/V_VC	0,1076	0,1180	0,1502	0,0721	0,1030	0,1071	0,1297	0,2218	0,3346	1,0361	3,6908	4,4226	0,1429	0,1685	0,2285	0,1519	0,1550	0,2090
	1-1/L_AG	0,1067	0,1178	0,1481	0,0698	0,0995	0,1058	0,1266	0,2157	0,3325	1,2035	3,8914	4,7696	0,1428	0,1676	0,2301	0,1502	0,1502	0,2003
	1-1/L_ET	0,1094	0,1210	0,1503	0,0729	0,0987	0,1038	0,1365	0,2202	0,3350	1,2414	4,0674	5,7328	0,1418	0,1657	0,2278	0,1515	0,1476	0,1795
	1-1/L_VC	0,1073	0,1172	0,1494	0,0709	0,1061	0,1028	0,1263	0,2162	0,3321	1,2021	4,0240	5,1204	0,1410	0,1689	0,2316	0,1488	0,1498	0,1953
	1-1/T_AG	0,1076	0,1190	0,1499	0,0712	0,0965	0,0964	0,1255	0,2105	0,3338	1,0599	4,1257	5,2680	0,1427	0,1678	0,2308	0,1483	0,1457	0,1881
	1-1/T_ET	0,1104	0,1206	0,1522	0,0733	0,0962	0,1016	0,1325	0,2137	0,3326	0,7564	3,7606	4,4730	0,1408	0,1666	0,2291	0,1467	0,1457	0,1929
	1-1/T_VC	0,1087	0,1178	0,1490	0,0729	0,1004	0,1019	0,1275	0,2116	0,3315	0,8275	3,9059	4,7535	0,1423	0,1683	0,2301	0,1520	0,1521	0,2047
1-1/V_AG	0,1074	0,1183	0,1480	0,0699	0,0917	0,1010	0,1323	0,2292	0,3381	1,1887	3,2068	5,2693	0,1422	0,1670	0,2327	0,1478	0,1428	0,1787	
1-1/V_ET	0,1081	0,1197	0,1547	0,0701	0,0937	0,0964	0,1340	0,2195	0,3314	1,2885	3,8525	6,5840	0,1487	0,1790	0,2339	0,1520	0,1506	0,1843	
1-1/V_VC	0,1084	0,1194	0,1479	0,0707	0,0933	0,0961	0,1366	0,2307	0,3390	1,3235	3,9611	6,5581	0,1422	0,1722	0,2368	0,1513	0,1492	0,1825	
Teste com a Rede	1/D_AG	0,1135	0,1242	0,1643	0,0716	0,0964	0,1056	0,1211	0,2215	0,3414	1,4673	3,8803	4,4875	0,1448	0,1637	0,2237	0,1462	0,1414	0,1789
	1/L_AG	0,1132	0,1225	0,1512	0,0722	0,1001	0,1097	0,1235	0,2135	0,3388	1,2640	3,6571	3,6793	0,1451	0,1664	0,2307	0,1466	0,1426	0,1853
	1/T_AG	0,1133	0,1219	0,1548	0,0722	0,0963	0,1042	0,1229	0,2107	0,3388	1,2634	3,6900	3,8968	0,1485	0,1720	0,2328	0,1457	0,1424	0,1714
	1/V_AG	0,1123	0,1229	0,1582	0,0729	0,1010	0,1090	0,1200	0,2222	0,3394	1,4997	3,9477	4,5945	0,1450	0,1697	0,2294	0,1459	0,1423	0,1823
	1-1/L_AG	0,1119	0,1222	0,1508	0,0756	0,1038	0,1095	0,1203	0,2111	0,3352	1,5693	3,9110	4,4412	0,1448	0,1671	0,2290	0,1495	0,1425	0,1798
	1-1/T_AG	0,1133	0,1233	0,1514	0,0746	0,1014	0,1045	0,1222	0,2097	0,3363	1,4272	4,0815	4,3558	0,1450	0,1691	0,2306	0,1475	0,1423	0,1811
	1-1/T_ET	0,1149	0,1233	0,1549	0,0727	0,0945	0,1086	0,1261	0,2120	0,3413	1,1530	3,7562	3,0157	0,1448	0,1667	0,2301	0,1465	0,1430	0,1728
	1-1/T_VC	0,1149	0,1208	0,1543	0,0719	0,1002	0,1110	0,1243	0,2084	0,3336	1,1272	3,5136	3,0771	0,1440	0,1680	0,2293	0,1470	0,1474	0,1792
	1-1/V_AG	0,1123	0,1221	0,1514	0,0720	0,0949	0,1042	0,1235	0,2216	0,3378	1,4738	3,5531	4,6584	0,1448	0,1663	0,2267	0,1458	0,1405	0,1706
	1-1/V_ET	0,1134	0,1231	0,1539	0,0720	0,0957	0,1068	0,1265	0,2196	0,3456	1,4503	3,6455	4,1233	0,1444	0,1654	0,2300	0,1476	0,1406	0,1642
1-1/V_VC	0,1135	0,1216	0,1545	0,0718	0,0968	0,1055	0,1233	0,2185	0,3387	1,4614	3,4616	4,2715	0,1450	0,1680	0,2303	0,1469	0,1405	0,1663	

Fonte: o autor

Tabela 5-4 – Análise de Acurácia dos Modelos Testados - RMSE

RMSE		Típico						Atípico A						Atípico B					
Teste	Tipo	Arterial			Expressa			Arterial			Expressa			Arterial			Expressa		
		LAG 1	LAG 5	LAG 15	LAG 1	LAG 5	LAG 15	LAG 1	LAG 5	LAG 15	LAG 1	LAG 5	LAG 15	LAG 1	LAG 5	LAG 15	LAG 1	LAG 5	LAG 15
Teste por Eixo	UN_ET	15,98	17,30	20,86	36,28	44,28	47,55	17,73	27,50	37,58	41,83	83,83	157,40	19,62	21,73	27,23	50,80	50,90	66,25
	UN_VC	15,95	17,28	20,45	35,75	46,20	48,43	17,53	27,75	37,55	43,28	84,10	157,33	19,57	21,98	27,40	50,98	51,68	66,75
	1/D_AG	16,17	17,34	20,55	35,45	45,30	46,83	17,48	27,67	38,17	44,25	86,55	161,50	19,67	21,80	27,32	51,13	51,93	67,25
	1/D_ET	15,97	17,30	20,85	35,95	43,98	48,08	17,85	27,55	37,58	43,63	87,23	160,53	19,62	21,83	27,33	51,18	51,35	67,48
	1/D_VC	16,01	17,33	20,50	35,70	44,83	47,40	17,53	27,90	37,87	43,95	86,28	160,53	19,62	21,78	27,17	51,30	51,80	67,58
	1/L_AG	16,17	17,38	20,63	35,68	45,20	47,95	17,48	27,67	38,23	42,73	83,90	157,78	19,63	21,73	27,30	51,28	51,58	67,35
	1/L_ET	16,03	17,32	20,68	35,98	44,55	47,73	17,63	27,40	37,60	42,43	84,08	157,10	19,80	21,70	27,18	50,65	50,23	64,00
	1/L_VC	16,04	17,36	20,53	35,95	45,20	45,90	17,57	27,82	37,85	42,18	83,88	159,55	19,60	21,72	27,12	50,88	51,70	65,60
	1/T_AG	16,23	17,43	20,84	35,45	45,80	47,53	17,57	27,58	38,48	42,88	83,68	158,13	19,70	21,73	27,42	50,83	51,58	66,93
	1/T_ET	16,00	17,31	20,86	35,73	44,70	48,88	17,80	27,62	37,77	41,75	82,10	154,88	19,58	21,77	27,35	50,40	50,55	65,98
	1/T_VC	16,05	17,43	20,53	36,05	45,30	47,78	17,52	27,93	37,92	41,63	82,75	156,88	19,67	21,65	27,00	50,78	51,45	66,80
	1/V_AG	15,88	17,23	20,40	35,35	44,18	49,13	17,07	27,37	37,15	45,15	85,55	156,13	19,52	22,25	27,58	51,30	49,90	65,80
	1/V_ET	15,97	17,27	20,85	35,80	44,50	47,60	17,77	27,38	37,33	42,18	83,70	157,18	19,60	21,82	27,25	50,48	50,65	65,30
	1/V_VC	15,95	17,23	20,37	35,65	45,90	46,50	17,45	27,55	37,23	44,25	85,55	160,38	19,55	22,15	27,55	51,50	51,75	66,88
	1-1/L_AG	15,89	17,14	19,78	35,23	45,18	45,85	16,80	26,60	36,87	44,88	87,58	162,15	19,57	21,75	27,05	52,03	50,60	65,38
	1-1/L_ET	16,14	17,39	20,21	35,70	45,05	46,03	17,93	27,55	37,65	44,80	91,98	167,25	19,55	21,60	27,00	51,90	49,98	63,05
	1-1/L_VC	15,93	17,13	20,05	35,55	46,30	44,30	17,02	26,83	37,03	43,73	87,75	163,75	19,55	21,87	27,23	51,38	50,50	63,88
	1-1/T_AG	16,03	17,28	19,97	35,63	43,45	42,28	16,73	25,83	36,63	45,45	91,48	167,73	19,65	21,88	27,15	51,28	49,58	63,30
1-1/T_ET	16,30	17,37	20,38	36,18	43,13	44,38	17,42	26,73	37,05	42,38	87,40	163,23	19,57	21,85	27,57	50,30	49,70	63,40	
1-1/T_VC	16,08	17,13	19,97	36,00	44,33	43,80	17,02	26,22	36,77	44,15	88,05	164,93	19,60	22,03	27,25	51,15	51,03	65,43	
1-1/V_AG	15,88	17,13	20,06	34,98	41,73	43,85	17,70	28,32	37,38	44,88	91,68	163,65	19,42	21,85	27,42	50,93	48,03	60,88	
1-1/V_ET	16,05	17,28	21,11	35,55	42,70	42,30	17,92	27,60	37,23	44,23	93,98	171,10	20,05	22,58	27,78	52,33	50,68	62,85	
1-1/V_vc	15,94	17,22	20,03	35,78	42,70	42,35	17,95	28,42	37,55	43,95	94,28	171,28	19,48	22,02	27,48	51,95	50,33	62,45	
Teste com a Rede	1/D_AG	16,85	18,17	22,34	35,53	43,50	45,70	16,63	28,08	40,05	48,80	92,43	164,20	20,15	22,20	28,50	48,95	47,65	61,08
	1/L_AG	16,72	17,71	20,32	35,73	45,18	47,83	16,60	26,47	37,38	46,75	87,88	157,80	20,00	21,80	27,25	49,05	48,30	61,88
	1/T_AG	16,76	17,83	21,08	35,78	44,28	46,95	16,52	26,15	37,33	47,43	92,18	163,90	20,40	22,47	27,60	48,90	47,75	59,65
	1/V_AG	16,77	18,15	21,53	36,15	44,93	46,80	16,40	27,63	38,73	48,98	91,53	164,58	20,13	22,55	28,37	49,05	47,55	61,55
	1-1/L_AG	16,72	17,86	20,28	36,85	46,35	47,63	16,28	26,10	36,95	48,30	89,95	161,68	20,10	21,87	27,12	49,98	47,78	61,28
	1-1/T_AG	16,86	18,02	20,28	36,73	45,78	46,05	16,42	25,82	36,82	48,78	90,85	164,93	20,17	22,07	27,30	49,58	47,68	61,58
	1-1/T_ET	16,76	17,72	20,48	35,83	44,40	47,95	17,05	27,32	38,28	47,53	95,53	163,28	19,88	21,35	27,17	49,30	47,45	58,50
	1-1/T_ET	16,93	17,93	20,94	35,98	43,75	48,05	16,78	26,28	37,32	47,65	93,65	163,78	20,03	21,85	27,33	48,83	47,88	59,78
	1-1/T_vc	16,80	17,71	20,78	35,73	44,43	47,40	16,93	27,57	38,03	47,73	95,20	162,23	19,92	21,71	27,60	49,08	47,50	58,70
	1-1/T_VC	16,96	17,78	21,02	35,65	44,78	48,48	16,85	26,23	37,05	48,30	92,05	164,15	20,13	22,20	27,57	49,10	48,95	61,58
1-1/V_AG	16,66	17,79	20,58	35,80	43,43	45,68	16,73	27,50	37,47	48,73	93,93	164,18	20,02	21,80	26,98	48,88	47,13	59,38	

Fonte: o autor

5.1.2 Estimação das Características do Tráfego com Dados em Painel

O estudo foi efetuado com a análise da estimação da velocidade pontual, buscando estimar a velocidade média nos sensores para o intervalo subsequente. A distribuição gaussiana foi utilizada em todos os modelos.

Devemos destacar que nesta etapa de estudo o eixo da Av. Brasil/Sumaré foi dividido em dois, sendo um com a Av. Brasil e a Av. Henrique Schaumann e outro com a Av. Sumaré e a Av. Paulo VI. Esta opção foi feita após análise mais detalhada das séries temporais, durante os estudos de STARIMA, onde verificou-se diferença no perfil de fluxo entre esses eixos.

A biblioteca “glmmTMB”(KAIN; BOLKER; MCCOY, 2015), disponível em linguagem R, foi utilizada nas análises dos modelos de Dados em Painel. Observa-se que a matriz de covariância padrão desta biblioteca é a não estruturada. No entanto, é possível o uso de outras estruturas, dentre as quais a de AR1.

Os modelos foram processados em computador com processador i7-8700 de 3,2GHZ da Intel, com 32Gb de memória RAM, sistema Windows 10 e disco SSD de 512 Gb. O tempo de processamento dos modelos nos ambientes de treino e teste foi aferido e será apresentado junto dos resultados na Seção 5.2.

Os modelos, que utilizam Dados em Painel, foram testados tanto na versão Longitudinal como hierárquica, do tipo HLM, de 2 e 3 níveis.

Os modelos HLM podem incluir estruturas de aninhamento dos dados ordenados de medidas repetidas ao longo do tempo, para cada indivíduo ou grupo de indivíduos. Nesta pesquisa verificaram-se duas possibilidades, indivíduo=Sensor e grupo=Eixo viário ou Classe da via, quando do aninhamento por “local” e indivíduo=Período e grupo=Dia da Semana ou Dia Útil no caso do aninhamento por “tempo”.

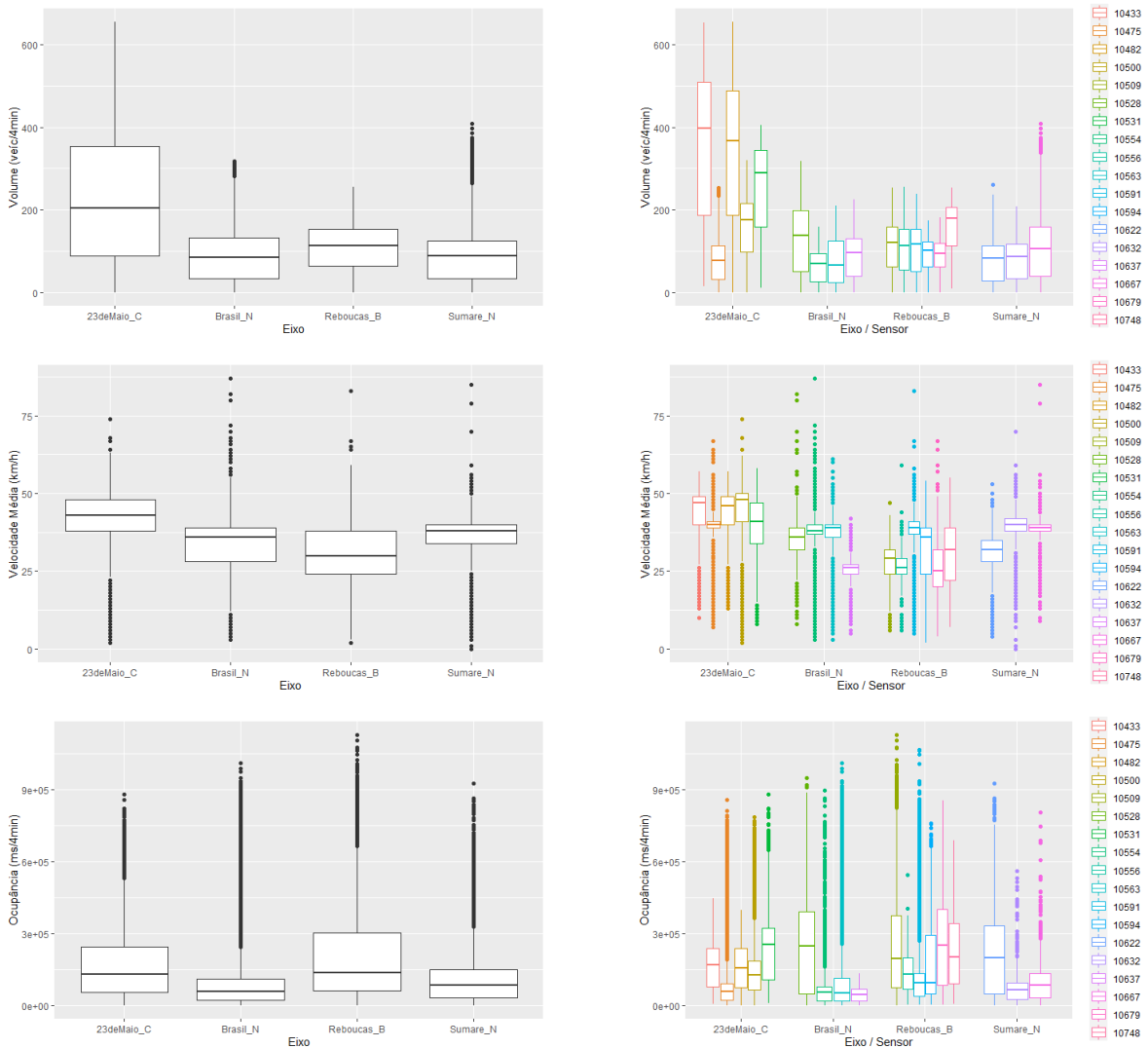
A Figura 5-5 e a Figura 5-6 mostram gráficos das três variáveis pontuais de estudo: Volume, Velocidade Média e Ocupação no Intervalo (ocupancia).

Na Figura 5-5 verifica-se os agrupamentos por local, sendo que à esquerda estão os gráficos relativos aos Eixos e à direita os que se referem a cada sensor (agrupados por eixo para melhor entendimento). É possível verificar que o Volume apresenta menor variação entre os eixos de vias arteriais, embora apresente alguma flutuação quando verificados os

sensores isoladamente. Com relação à Velocidade média e à Ocupação no Intervalo (ocupancia), ocorrem diferenças claras tanto entre os eixos como entre os sensores.

Da existência destas variações, é possível avaliar, visualmente, a possibilidade de agrupamento dos sensores por local. Disto resultou que os cenários testados por local incluíram tanto o nível Eixo como o nível Sensor.

Figura 5-5 – Dados aninhados por local - Eixo e Sensor

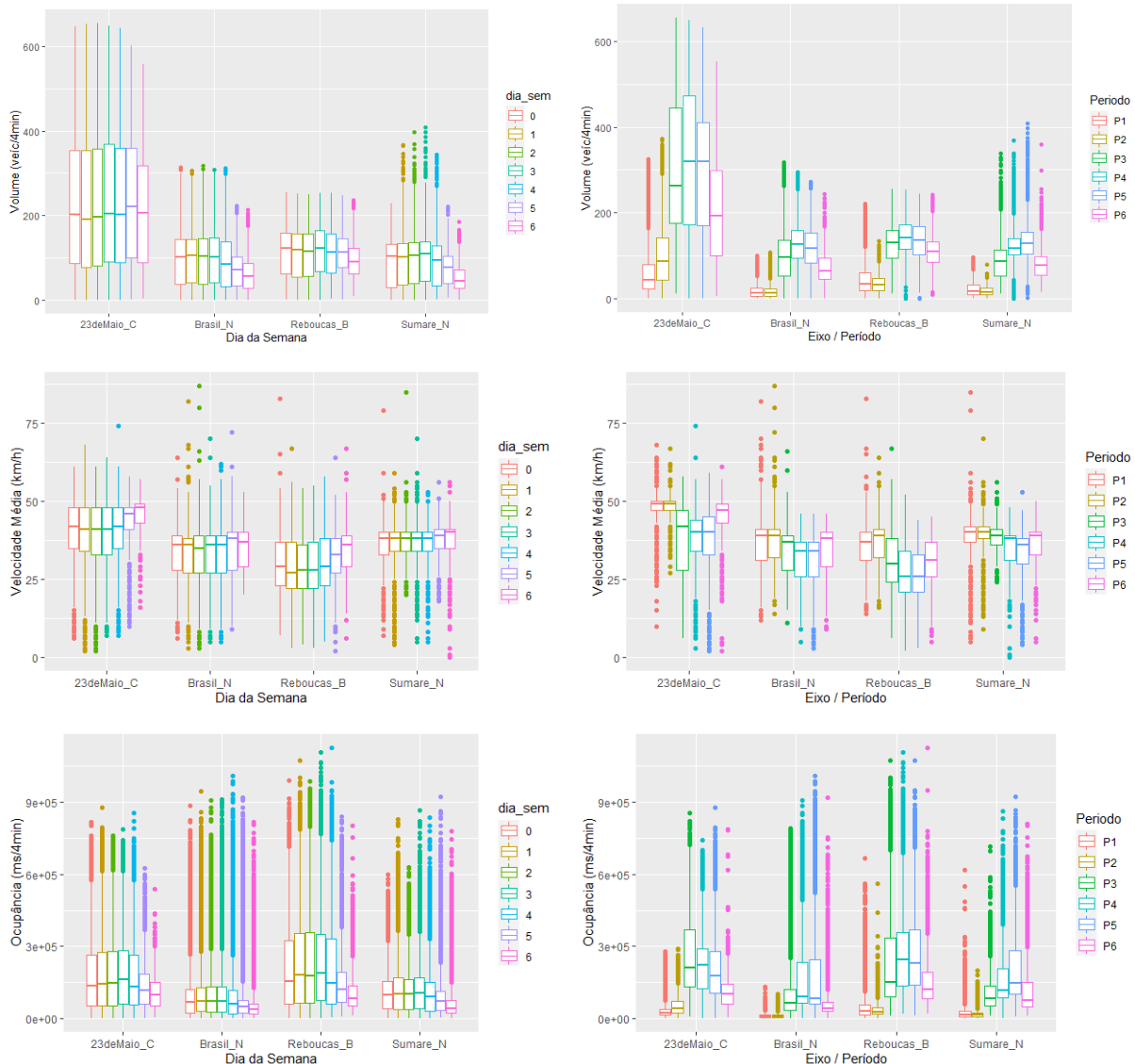


Obs. Nos gráficos à direita, cada sensor está apresentado com uma cor.
 Fonte: o autor

Considerando agora os agrupamentos baseados no tempo, apresentados na Figura 5-6 divididos por eixo de estudo, verifica-se que ocorrem diferenças nas medidas quando avaliados os períodos do dia. Porém, as diferenças são pequenas entre os dias da semana, para um mesmo eixo de vias.

Da análise visual da Figura 5-6, se verifica que a variação existente entre períodos sugere o aninhamento dos dados por aspecto temporal. O oposto ocorre em relação aos dias da semana, nos quais ocorre uma variação maior nos finais de semana. Assim, verificamos a possibilidade de testar a hierarquização temporal tanto pelos dias da semana como de modo mais simplificado, separando-se apenas os dias úteis dos finais de semana.

Figura 5-6 – Dados aninhados por tempo – Dia da Semana e Período do dia



Obs. Nos gráficos, cada período ou dia da semana está apresentado com uma cor.

Fonte: o autor

Uma vez que as duas formas de hierarquização se mostraram possíveis, o estudo foi iniciado com os modelos em paralelo, a fim de identificar qual abordagem resultaria em melhores resultados. Seguiu-se então ao processo proposto no item 3.1.3, com a análise em

separado das variáveis de efeito fixo e de efeito aleatório, concluindo-se o processo com a montagem de modelos hierárquicos com efeito fixo e aleatório.

Os modelos baseados em Dados em Painel foram avaliados durante a fase de treino através do índice AIC e as predições feitas no ambiente de teste, tiveram a análise de acurácia baseada nos medidores MAPE e RMSE.

5.2 DISCUSSÃO DOS RESULTADOS

Nesta seção são apresentados os resultados dos modelos elaborados. A análise dos resultados dos estudos sobre matrizes do modelo STARIMA está apresentada em separado, tendo em vista a abrangência do estudo que se fez necessário para permitir o passo seguinte, relativo à estimação com esse modelo. Em seguida, são apresentados e discutidos os resultados obtidos com os modelos baseados em Regressão com Dados em Painel. Por fim, são comparados os resultados de estimação obtidos com STARIMA e Dados em Painel, com base na estimação da velocidade pontual.

5.2.1 Estimação dos modelos com modelo STARIMA

O modelo STARIMA é baseado no uso de matrizes de peso. Sua aplicação exige uma boa compreensão de suas características e dos efeitos decorrentes. A seguir são apresentados os resultados do estudo específico sobre características das matrizes, aplicados sobre a variável quantitativa volume (fluxo), cuja aplicação ocorreu num período de 2 horas do dia. Este item 5.2.1 é concluído com os modelos STARIMA aplicados para a variável quantitativa velocidade, no banco de testes, nas 24 horas do dia, visando à comparação com os modelos de Dados em Painel.

5.2.1.1 Vizinhança

Os estudos buscaram avaliar e comparar a utilização das matrizes VC e LP, que são os tipos já utilizados nos estudos de ST-STTF, baseados no modelo STARIMA, com a aplicação da matriz AG, utilizada em outras áreas de estudo, mas ainda não aplicada na área de previsão de tráfego com STARIMA.

Os valores apresentados na Tabela 5-5 resumem os resultados anteriores da Tabela 5-3 e da Tabela 5-4, mostra a porcentagem de vezes em que um modelo apresentou a melhor acurácia para determinado sensor. Verifica-se, com base no MAPE, que o tipo AG se equipara aos outros dois quando nos estudos dos corredores em separado. Nos testes da Rede, este foi o modelo que apresentou melhor resultado em 42% dos cenários. Com base no RMSE, o tipo AG é o melhor em 58% dos cenários com toda a rede e em 44% dos cenários com a via arterial.

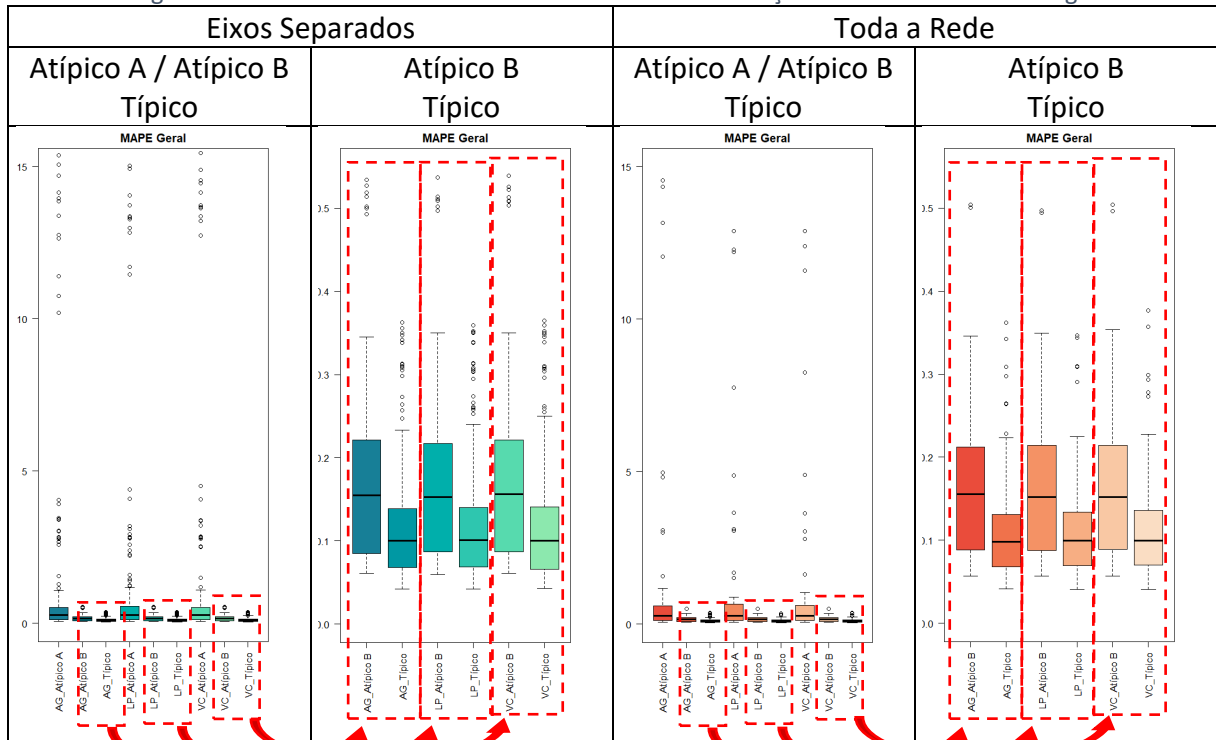
Tabela 5-5 – Porcentagem de vezes em que o modelo obteve melhor acurácia de acordo com a Vizinhança

Índice de Acurácia	Vizinhança	Expressa	Arterial	Rede
MAPE	AG	43%	33%	42%
	VC	13%	43%	33%
	LP	44%	25%	25%
RMSE	AG	39%	44%	58%
	VC	17%	33%	29%
	LP	46%	31%	17%

Obs. No caso de empate os dois modelos foram considerados
 Fonte: o autor

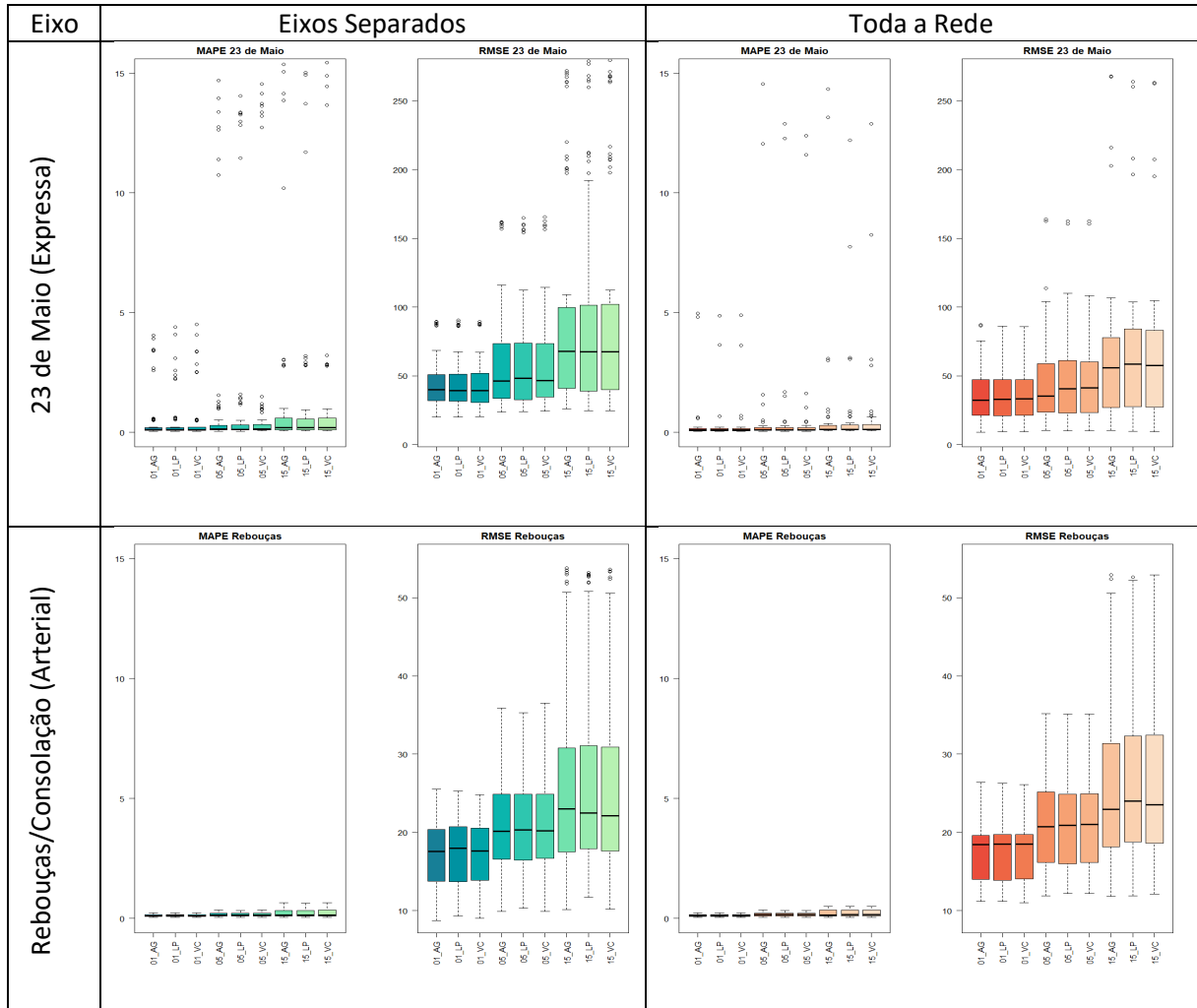
Com relação à acurácia, é possível verificar na Figura 5-7 que considera os horizontes de previsão, vê-se que o modelo AG, de modo geral, apresenta dispersão ligeiramente maior que os demais. Na Figura 5-8, vê-se que o modelo AG apresenta dispersão semelhante aos demais.

Figura 5-7 – Precisão da Previsão de acordo com a Vizinhança e Característica do Tráfego



Fonte: o autor

Figura 5-8 – Acurácia de Previsão de acordo com o tipo de Vizinhança e Horizonte de Previsão



Fonte: o autor

Desta forma, é possível concluir que o modelo AG, até agora inédito em aplicações de ST-STTF, pode ser utilizado nos estudos de ST-STTF com o STARIMA, simplificando assim o modelo e reduzindo a demanda computacional necessária.

5.2.1.2 Pesos

Comparando-se os modelos assimétricos menos utilizados nos estudos com STARIMA, com os simétricos (1/T, 1/V, 1/D e 1/L) que são mais comuns, verifica-se superioridade dos modelos assimétricos (1-1/V, 1-1/T e 1-1/L) tanto para a rede como para as vias em separado. Novamente com base na Tabela 5-3 e Tabela 5-4 foi elaborada a Tabela 5-6, que compara os tipos de peso testados, verificando-se que na maioria das vezes os modelos assimétricos são

melhores, com destaque ao modelo 1-1/V que aparece mais vezes entre os de melhores resultados.

Tabela 5-6 – Comparação da Acurácia por Tipo de Peso (MAPE)

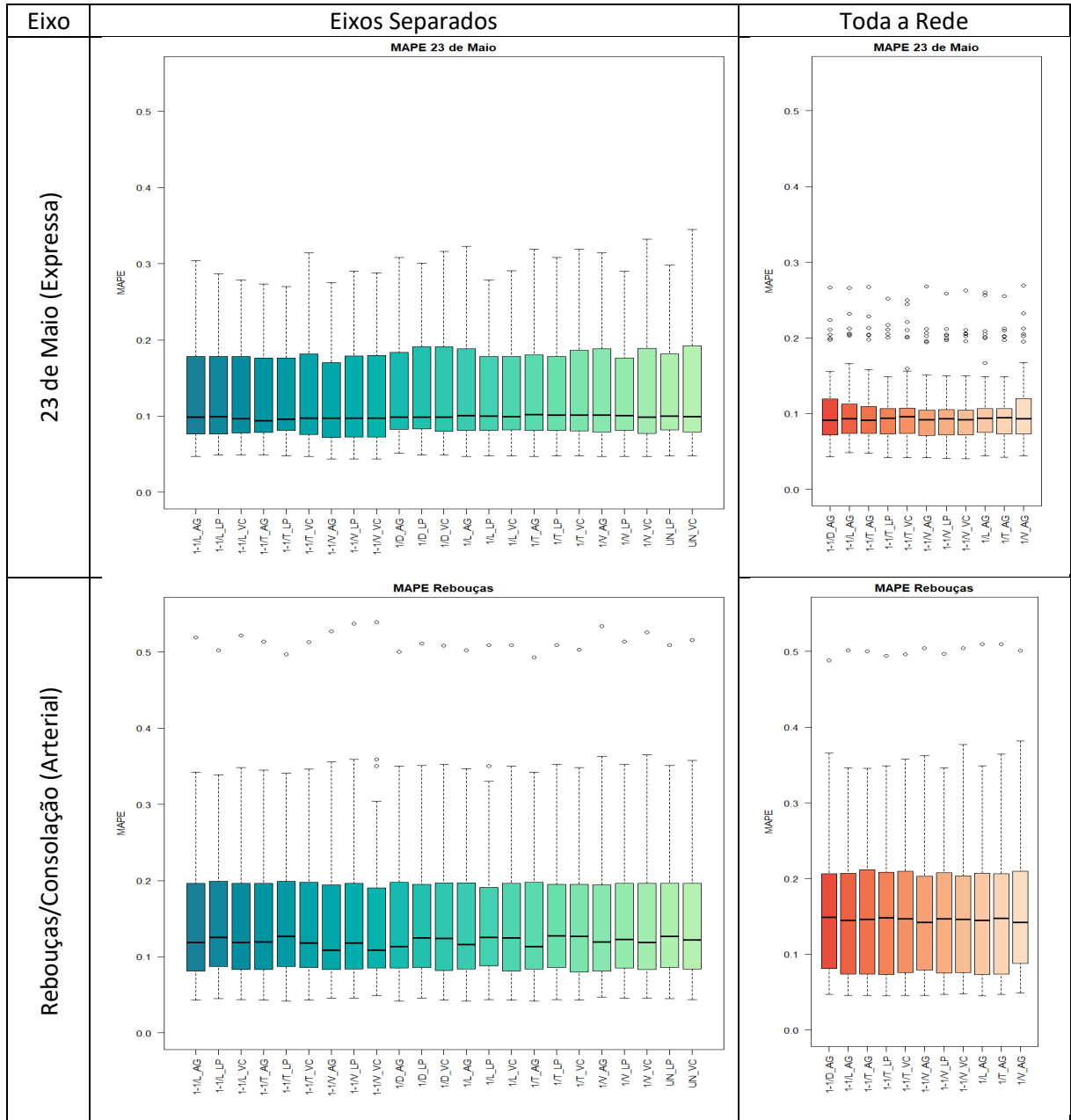
Tipo de Peso	Eixos Separados			Rede			Total Geral		
	1º Melhor	2º Melhor	1º + 2º	1º Melhor	2º Melhor	1º + 2º	1º Melhor	2º Melhor	1º + 2º
1-1/V	38,9%	15,0%	26,3%	25,0%	52,6%	38,5%	31,6%	33,3%	32,5%
1-1/T	22,2%	30,0%	26,3%	35,0%	26,3%	30,8%	28,9%	28,2%	28,6%
1-1/L	16,7%	20,0%	18,4%	10,0%	10,5%	10,3%	13,2%	15,4%	14,3%
1/T	16,7%	15,0%	15,8%	10,0%		5,1%	13,2%	7,7%	10,4%
1/V	5,6%	10,0%	7,9%	5,0%	5,3%	5,1%	5,3%	7,7%	6,5%
1/D		5,0%	2,6%	15,0%		7,7%	7,9%	2,6%	5,2%
1/L		5,0%	2,6%		5,3%	2,6%		5,1%	2,6%

Fonte: o autor

A Figura 5-9 apresenta os resultados obtidos em cada um dos dois corredores para todos os 34 modelos testados, incluindo os testes dos dias Típico e Atípico B, onde se verifica que os modelos assimétricos apresentaram menor dispersão do que os simétricos.

Assim, embora menos utilizados, os modelos assimétricos se apresentaram mais adequados às previsões em ambiente urbano, uma vez que incorporam a diferença da influência entre o tráfego que chega e o que sai de um ponto. Como discutido na seção 3.2, o fluxo mais denso, com menor velocidade e, portanto, com maior tempo de percurso à jusante do ponto de medição, pode restringir o fluxo no local. Já, se a situação ocorre à montante o impacto pode ser o inverso, já que ocorre certa restrição na chegada dos veículos no ponto de medição

Figura 5-9- Acurácia de previsão dos pesos testados em separado e na rede



Fonte: o autor

5.2.1.3 Abrangência

A comparação dos modelos, nos 3 tipos de abrangência, foi feita com base na diferença de acurácia obtida nos dois Eixos testados.

De forma geral, vemos nas Tabela 5-3 e Tabela 5-4 que os modelos que incluíram toda a rede apresentaram resultados um pouco piores para as previsões do que os modelos construídos em separado, confirmando estudos anteriores onde os resultados de modelos de

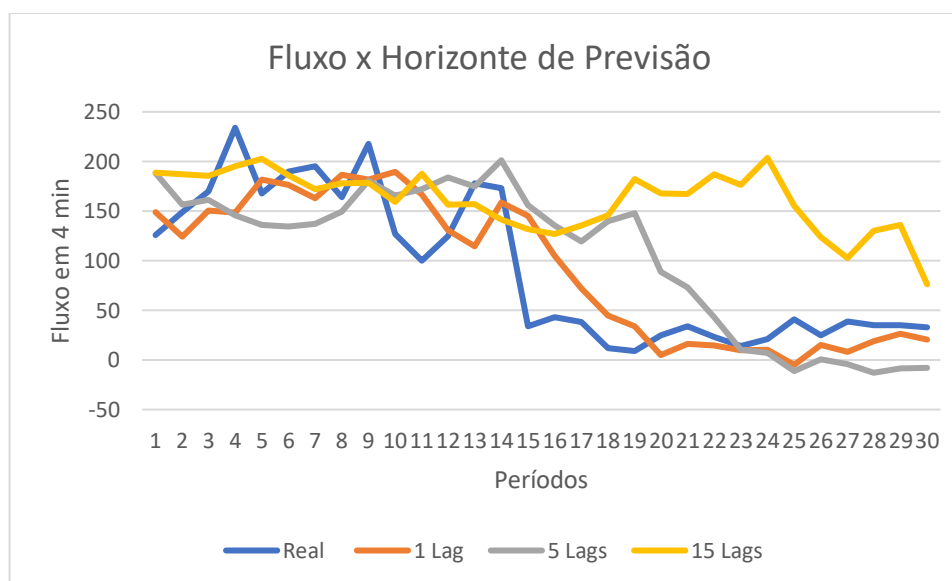
STARIMA desenvolvidos para cada sensor superaram em mais de 10% os resultados do modelo que considerou a rede como um todo (CHENG et al., 2014). Isto pode estar correlacionado com as diferenças de fluxo entre os corredores.

5.2.1.4 Horizonte de previsão

Com relação ao Horizonte de previsão, como era esperado, à partir da literatura revisada (ERMAGUN; LEVINSON, 2018; LAÑA et al., 2018; PAVLYUK, 2020), verificou-se que os modelos com horizontes maiores apresentaram menor acurácia. Nos dias típicos a acurácia das previsões permitiu erros medidos pelo MAPE a partir de 7% na via expressa e 10% na via arterial para o horizonte de 1 passo (4 min). Para os horizontes de 5 e 15 passos esses valores se ampliam, atingindo índices de até 11% na via expressa e de 24% na via arterial, nos dias típicos.

Nos dias atípicos a perda de acurácia também se intensifica nas previsões com horizontes mais distantes, uma vez que o modelo utiliza dados passados para prever o futuro e o impacto de variações bruscas demoram mais a impactar o modelo, quando feitas previsões com horizontes maiores, como podemos observar nos gráficos da Figura 5-10.

Figura 5-10 – Gráfico da previsão com vários horizontes



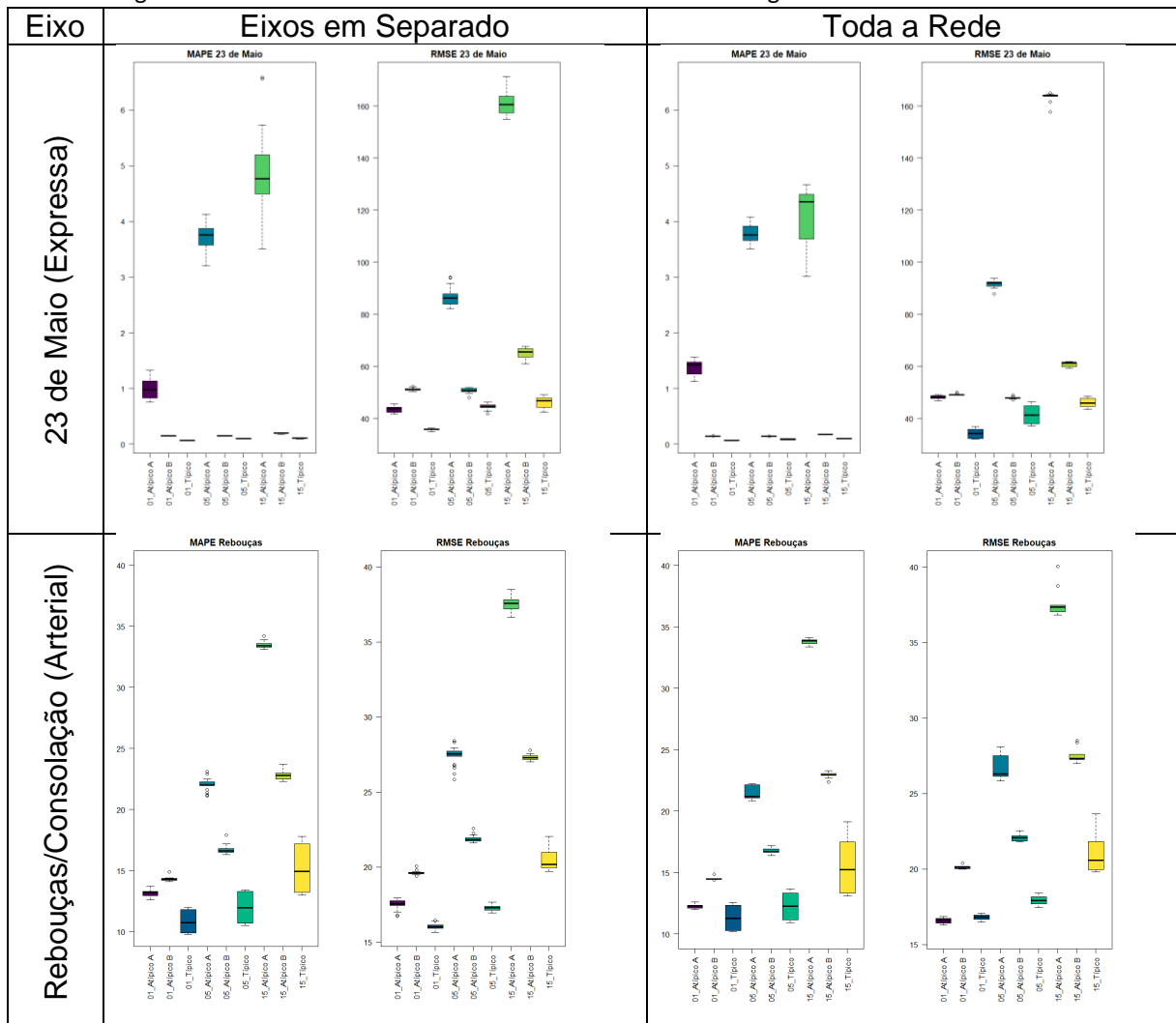
Fonte: o autor

5.2.1.5 Característica do tráfego no período

O modelo STARIMA mostrou-se capaz de prever o fluxo em dias típicos, com erros abaixo de 10% para horizontes de 1 passo, e em torno de 15% para 15 passos à frente. Para previsões em situações atípicas decorrentes de feriado (Atípico B) verificou-se erros maiores, o que indica a recomendação de elaborar modelos específicos para dias que apresentam perfil de fluxo diferente da média geral, Figura 5-11.

Para os dias atípicos, ou com incidentes, que impliquem na redução brusca da série histórica, como o Atípico A, especialmente se ocorrendo em apenas parte dos locais da rede, o modelo apresenta acurácia razoável para 1 passo. Entretanto, acima disto as previsões tendem a perder precisão, conforme se verifica através do índice MAPE, onde ocorre perda acentuada na acurácia nos horizontes de 5 e 15 passos à frente do dia Atípico-A.

Figura 5-11 –Acurácia de Acordo com a Característica do Tráfego e o Horizonte de Previsão



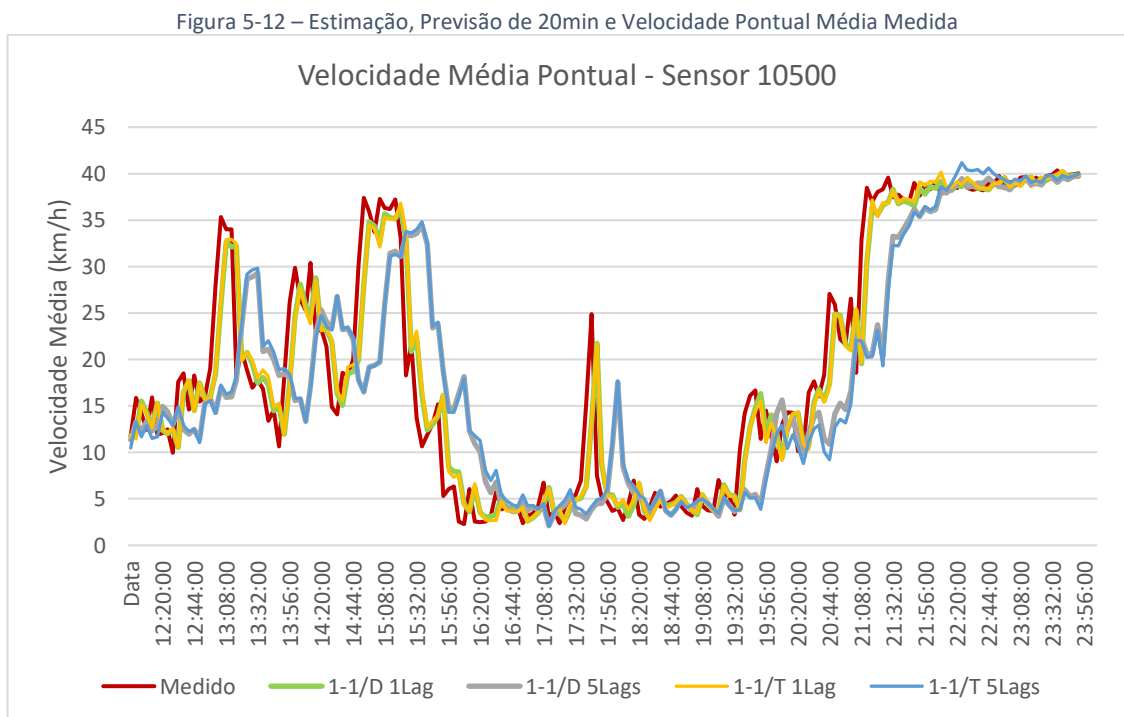
Fonte: o autor

5.2.1.6 Modelos STARIMA para a característica velocidade pontual

Os modelos desenvolvidos com os dados de treino foram testados com a aplicação da matriz do tipo 1/D, que não se altera de acordo com as flutuações do tráfego, os testes foram complementados com a matriz 1-1/T, que considera as diferenças de tempo de percurso.

Os testes utilizaram-se das matrizes elaboradas para o estudo de matrizes. Salienta-se que os testes para as 24 horas do dia apresentaram resultados bastante satisfatórios, ainda que se considerando que este modelo foi desenvolvido para o período das 12h às 16h. Esta aplicação estendida já havia sido parcialmente verificada com sucesso nos testes com as matrizes, ao se fazer a previsão até às 17h. A criação de modelos específicos para cada período do dia é possível e poderá ser explorada em estudos futuros.

Os modelos assimétricos e com vizinhança agrupada (AG) se mostraram superiores nos testes apresentados nas seções anteriores. A Figura 5-12 apresenta uma comparação entre as estimações (4 min) e previsões de 20 min, feitas com os modelos 1-1/T (assimétrico) e 1/D (simétrico), para o sensor 10500, num período de 12 horas do dia 20/03/2022, com tráfego atípico devido a chuvas.



Fonte: O autor

Verifica-se que tanto as estimações (1 Lag) como as previsões para 20 min (5 Lags), com cada um dos tipos de pesos, produzem resultados muito semelhantes na maior parte do tempo. As diferenças a favor do modelo 1-1/T, que considera o tempo de percurso, em relação ao 1/D, que considera a distância do percurso, ocorrem especialmente nos momentos de mudanças mais bruscas na velocidade medida.

Também é possível constatar que os modelos de estimação produzem resultados que se aproximam dos valores medidos e que os modelos de previsão de 20 min indicam resultados menos precisos e indicam certa defasagem na assimilação de variações maiores dos valores medidos. A Tabela 5-7 apresenta o desempenho dos modelos avaliados na estimação da velocidade, para os dias 20 e 21/03, separando-se o período das 15 às 21h que estiveram sob impacto de chuvas intensas. Os dados confirmam que, diferentemente do que ocorreu nos testes com a variável “volume”, nestes testes com a variável velocidade os dois modelos mostram desempenho similar, com ligeira vantagem para o modelo 1-1/T.

Tabela 5-7 – Desempenho dos modelos STARIMA na estimação da velocidade pontual

Medidor de Acurácia	Situação	STARIMA 1-1/T	STARIMA 1/D
MAPE	Chuvas Fortes (Atípico)	0,1558	0,1569
	Exceto Chuvas Fortes (Típico)	0,0696	0,0696
	Geral	0,0804	0,805
RMSE	Chuvas Fortes (Atípico)	3,708	3,720
	Exceto Chuvas Fortes (Típico)	3,022	3,022
	Geral	3,116	3,118

Fonte: o autor

Os resultados das estimações para 24 horas do dia, com base na matriz do tipo 1/D, são apresentados na Figura 5-13. Verifica-se que, para as 3 variáveis de tráfego avaliadas, Volume (Fluxo), Ocupação e Velocidade Pontual, os valores estimados, em amarelo, acompanham e se aproximam dos valores reais apresentados em azul. Essa proximidade, ocorre também em períodos com variações atípicas, como por exemplo, entre os intervalos 1700 e 1750, que correspondem ao período atípico com chuvas fortes no dia 20/03/18. Neste mesmo intervalo de tempo, com chuvas fortes, verifica-se que os valores da ocupação são muito superiores aos valores dos demais intervalos e as velocidades ficam abaixo de 5km/h.

Figura 5-13 – Estimação das Características de Tráfego – STARIMA 1/D



Fonte: o autor

Assim, embora com alguma perda de precisão, os modelos que consideram a distância entre os pontos podem ser utilizados, eliminando do processo a necessidade de determinação dos tempos de percurso, que é trabalhosa e apresenta diversas falhas na obtenção dos dados.

5.2.2 Estimação com modelo de Dados em Painel

A montagem dos modelos de estimação da velocidade pontual, baseados em Dados em Painel, foi feita analisando-se cada parte do modelo e por fim unificando as melhores características em modelos hierárquicos com variáveis aleatórias e fixas.

5.2.2.1 Modelos hierárquicos nulos

Para análise da estrutura aninhada, como parte do estudo inicial, foram elaborados os modelos nulos, quando não ocorre a inclusão de variáveis explicativas, para os dois tipos de aninhamento avaliados. Foram testados 11 modelos nulos, conforme Tabela 5-8.

Tabela 5-8 – Desempenho dos modelos Nulos

Modelo	Descrição dos modelo	AIC	Sigma2	MAPE	RMSE	Tempo de treino (s)	Tempo de teste (s)
N1	Nível 2 = Sensor e Nível 3=Eixo	803704,5	44,1	21,64%	7,397	3,33	2,86
N1N	Nível 2 = Sensor	803708,7	44,1	21,64%	7,397	2,69	2,14
N1E	Nível 2 = Eixo	839036,7	59,0	25,37%	8,201	2,75	2,11
N1C	Nível 2 = Sensor e Nível 3=Classe	803243,5	44,1	21,64%	7,397	3,02	2,70
N1_ARE	N1 com ARE	730631,9	23,1	21,82%	7,446	43,03	1480,22
N1_ARN	N1 com ARN	585442,5	3,6	21,59%	7,372	138,66	2199,13
N1N_ARN	N1N com ARN	585446,7	3,6	21,59%	7,373	125,11	2080,55
N2	Nível 2 = Período e Nível 3= Dia da Semana	855916,7	67,8	25,87%	8,538	4,07	3,04
N2_AR	N2 com ARN	849998,0	64,2	26,08%	8,552	54,30	1249,87
N3	Nível 2 = Período e Nível 3= Dia Útil	856557,3	68,2	25,83%	8,505	3,76	3,00
N3_AR	N3 com ARN	850470,5	64,5	26,02%	8,519	54,76	1260,60

Obs. ARE = Autorregressão com base no Eixo viário, ARN = Autorregressão com base no Sensor, Sigma2 = variância residual

Fonte: o autor

Para os modelos nulos com base no local, verificamos que o modelo N1N, que agrupa os dados apenas pelo Sensor, apresentou acurácia igual aos modelos onde se incluiu mais um nível, seja com a utilização do Eixo (N1) seja com o uso da Classe das vias (N1C). A utilização de estruturas de covariância autorregressivas deve tomar como base uma variável hierárquica. Comparando-se o uso do Sensor (N1_ARN) e do Eixo (N1_ARE), verificamos que o modelo com o sensor apresentava acurácia maior. Novamente o modelo autorregressivo com a variável Eixo no nível hierárquico 3 (N1_ARN) mostrou a mesma acurácia do modelo sem esse nível (N1N_ARN).

A comparação entre os modelos com base local, sem e com autorregressão, N1 e N1_ARN respectivamente, estão apresentados na Figura 5-14. Nos gráficos abscissa indica os valores reais e a ordenada corresponde aos valores estimados. Assim, a reta diagonal tracejada, sobreposta aos gráficos, corresponde ao que seria uma perfeita estimação de todos os dados. As linhas coloridas correspondem à estimação relativa à cada sensor, sendo a legenda apresentada à direita das imagens.

Verifica-se que, no cenário de treino aninhado por Local (Figura 5-14) que o uso da correlação por autorregressão apresentou uma melhora significativa no parâmetro de comparação AIC. Essa melhora é perceptível no gráfico superior direito da Figura 5-14, no qual verifica-se que as linhas coloridas, que correspondem à média das estimações para cada sensor, se aproximam mais da diagonal. No entanto, o resultado com os dados de teste não seguiu na mesma direção, como se percebe através dos dois gráficos inferiores da mesma Figura 5-14, que são semelhantes e com linhas quase horizontais, mostrando pouca acurácia do modelo, conforme constatado pelos valores de acurácia MAPE e RMSE na Tabela 5-8. Nesta direção, constata-se que a autorregressão causou uma espécie de overfitting de modo que o modelo se adaptou bem aos dados de treino, e não tão bem aos dados de teste.

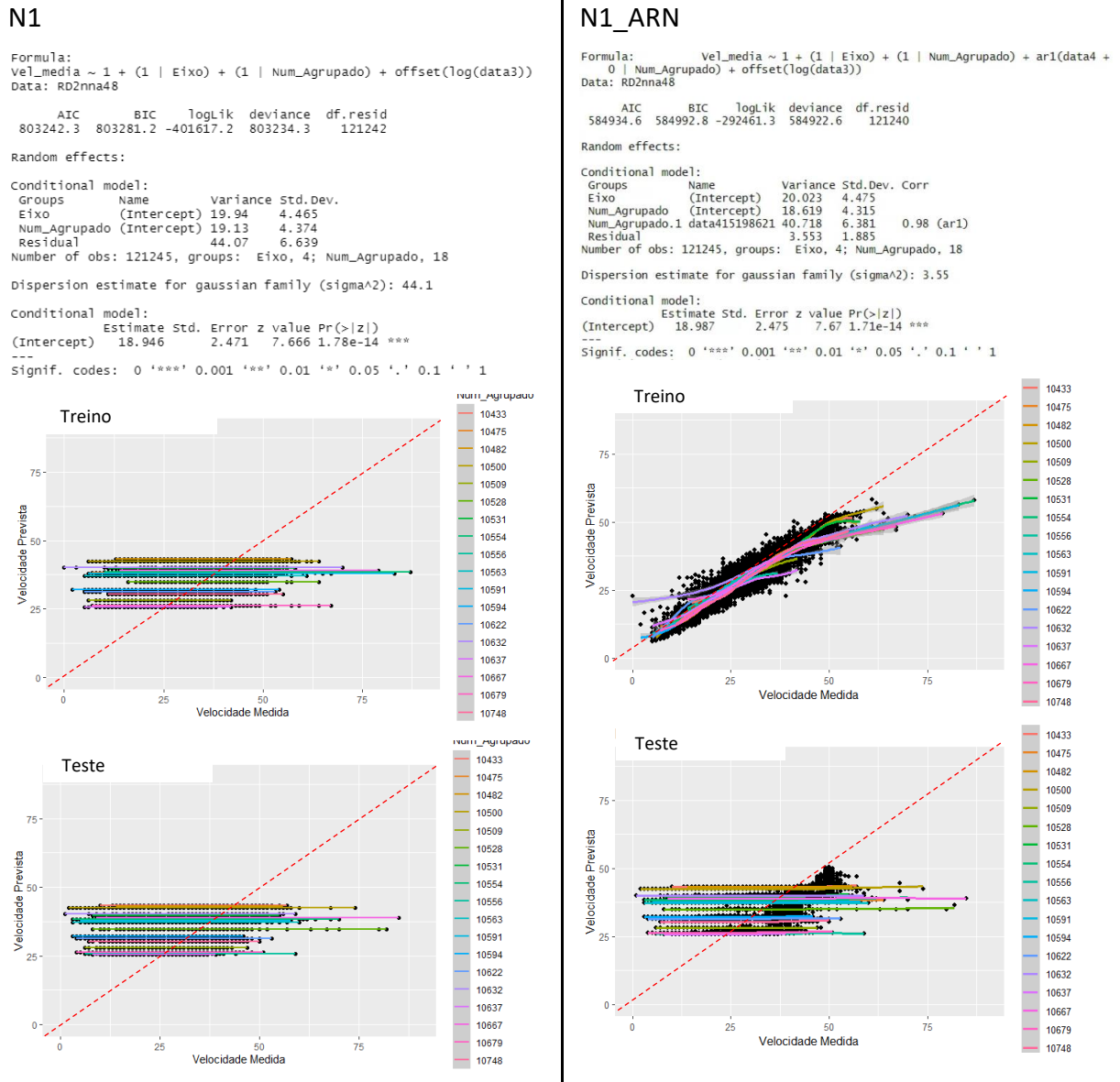
Os modelos utilizados, na construção de cada cenário apresentado, estão apresentados na parte superior das figuras. Na descrição são indicados: o tipo de distribuição utilizada (gaussian = gaussiana); a fórmula testada, na qual verifica-se que a variável dependente é a velocidade média (Vel_media); os aninhamentos como por exemplo, eixo (1|Eixo) ou Sensor (1|Num_Agrupado); a base de autorregressão; os índices de qualidade do modelo: AIC, BIC, LogLikelihood e desvios e ainda a análise das variáveis.

Destaca-se que o tempo de processamento com uso de autorregressão é muito superior. Na etapa de teste os modelos autorregressivos levaram mais de vinte minutos para efetuar a

previsão dos onze dias de teste, enquanto os modelos com covariância não estruturada necessitaram de apenas três segundos. Esse fato é relevante e negativo, já que nosso estudo visa a utilização de processamento em tempo próximo ao real.

Assim, dos modelos hierárquicos com base no local, optamos pelo prosseguimento dos estudos com o modelo que considerava apenas o Sensor (N1N), sem autorregressão.

Figura 5-14 – Modelo Nulo Aninhado por Local



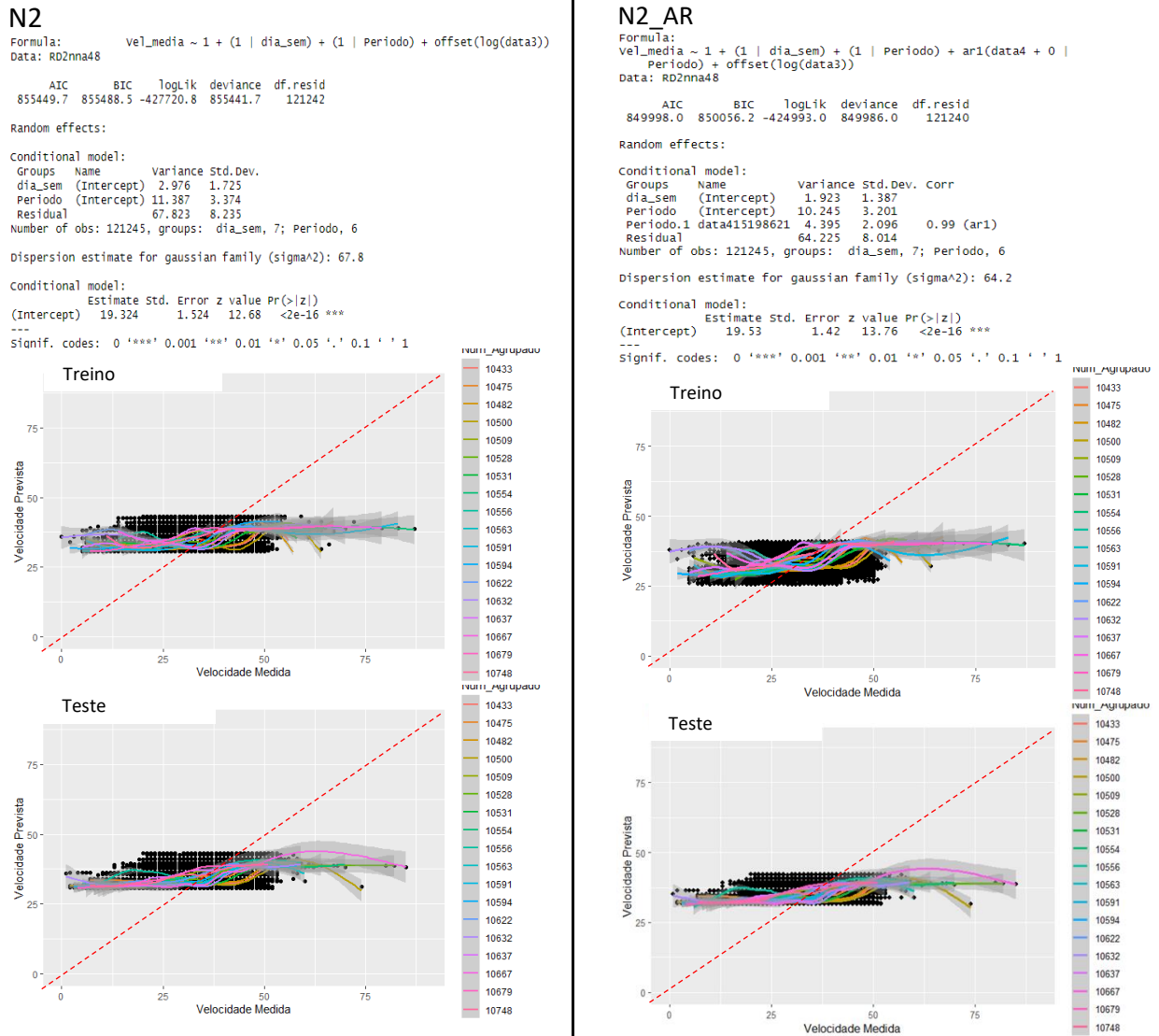
Obs. Nas abscissas estão os valores reais e nas ordenadas os valores estimados. À esquerda o modelo com correlação não estruturada e à direita com correlação AR1

Fonte: o autor

De modo similar, testamos os modelos nulos com base no tempo. Na Figura 5-15, que trata do Modelo Nulo aninhado por tempo com o Dia da Semana no nível 2, verificamos pelo índice AIC apenas uma pequena melhora no modelo após a alteração da forma de correlação

para o modo autorregressivo. Isso se confirma quando examinados os valores de Variância dos Efeitos Aleatórios, onde se verifica que a variável “Dia da Semana” apresenta variância muito inferior à dos “Períodos”. Quanto ao modelo N3, que considera a separação baseada nos dias úteis, verificamos características similares às do modelo N2 baseado nos dias da semana.

Figura 5-15 – Modelo Nulo Aninhado por Tempo



Obs. Nas abscissas estão os valores reais e nas ordenadas os valores estimados. À esquerda o modelo com correlação não estruturada e à direita com correlação AR1

Fonte: o autor

Os gráficos dos modelos aninhados pelo tempo apresentam certa flutuação, todavia isto não significa um modelo melhor que os modelos com base no local. A comparação entre modelos nulos, feita com base nos índices de acurácia AIC, MAPE e RMSE, mostra que o Modelo Nulo por Local é superior ao Modelo Nulo por Tempo, sendo selecionado para os

testes seguintes. Outro fator comparativo que confirma essa conclusão é a Variância Residual (σ^2), que no aninhamento por Local (N1N) apresentou o valor de 44,1 e para o aninhamento por tempo (N2) foi superior a 64.

Assim, dentre os 11 modelos nulos testados optamos pelo modelo N1N devido sua melhor acurácia em relação aos demais modelos. A opção pelo modelo sem autorregressão deveu-se à pequena diferença de acurácia e ao tempo de processamento menor.

5.2.2.2 Análise das variáveis aleatórias

Seguindo com o estudo, foi verificada a inclusão de variáveis nos efeitos aleatórios no modelo N1N, tendo sido testados os 5 modelos que estão apresentados na Tabela 5-9, em conjunto com o modelo nulo N1N (sem variáveis aleatórias) para facilitar a comparação.

Tabela 5-9 – Desempenho dos modelos com efeitos aleatórios

Modelo	Descrição dos modelo	AIC	Sigma2	MAPE	RMSE	Tempo de treino (s)	Tempo de teste (s)
N1N	Modelo Nulo sem efeitos aleatórios	803708,7	44,1	21,64%	7,397	2,69	2,14
N1N_A1	N1N com Dia Útil e Período	724211,4	22,8	15,94%	7,397	15,70	4,18
N1N_A2	N1N com Período	744092,9	26,9	16,67%	6,017	6,21	2,40
N1N_A3	N1N com Dia da Semana e Período	719638,9	21,9	15,93%	5,765	14,24	2,86
N1N_A4	N1N com Dia da Semana	789020,9	39,1	20,54%	7,189	6,62	2,55
N1N_A5	N1N com Dia Útil	791506,6	40,0	20,50%	7,158	3,33	2,05

Obs. ARE = Autorregressão com base no Eixo viário, ARN = Autorregressão com base no Sensor, σ^2 = variância residual

Fonte: o autor

Da tabela acima podemos constatar que a inclusão de efeitos aleatórios melhorou os modelos, tanto no ambiente de treino, medido pelo AIC, como no de testes onde a comparação é feita pelo MAPE e pelo RMSE. Os modelos N1N_A1, N1N_A2 e N1N_A3 se destacaram, sendo o modelo N1N_A3 o que apresentou melhor acurácia.

A Figura 5-16 apresenta os modelos N1N_A2 e N1N_A3, sendo de se notar que as linhas coloridas, que indicam o comportamento do modelo para cada sensor, passaram a apresentar uma configuração mais inclinada, aproximando-se da linha diagonal. Portanto, a comparação

destes gráficos com os da Figura 5-14 e da Figura 5-15, que indicam linhas próximas à horizontal, permite a confirmação de que a inclusão do efeito aleatório melhorou o modelo.

Figura 5-16 – Modelos Aninhados por Local com Efeitos Aleatórios

N1N_A2

Formula: $Vel_media \sim 1 + (Periodo | Num_Agrupado) + offset(\log(data3))$
 Data: RD2nna48

	AIC	BIC	logLik	deviance	df.resid
	744092.9	744316.2	-372023.5	744046.9	121223

Random effects:

Conditional model:

Groups	Name	Variance	Std.Dev.	Corr
Num_Agrupado	(Intercept)	102.778	10.138	
	PeriodoP2	1.022	1.011	-0.07
	PeriodoP3	58.675	7.660	-0.78 -0.07
	PeriodoP4	69.480	8.335	-0.79 -0.29 0.87
	PeriodoP5	77.971	8.830	-0.78 -0.37 0.69 0.92
	PeriodoP6	13.842	3.720	-0.60 -0.43 0.64 0.88 0.91
	Residual	26.945	5.191	

Number of obs: 121245, groups: Num_Agrupado, 18

Dispersion estimate for gaussian family (sigma^2): 26.9

Conditional model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	16.084	1.826	8.808	<2e-16 ***

N1N_A3

Formula: $Vel_media \sim 1 + (dia_sem + Periodo | Num_Agrupado) + offset(\log(data3))$
 Data: RD2nna48

	AIC	BIC	loglik	deviance	df.resid
	719638.9	720415.4	-359739.5	719478.9	121166

Random effects:

Conditional model:

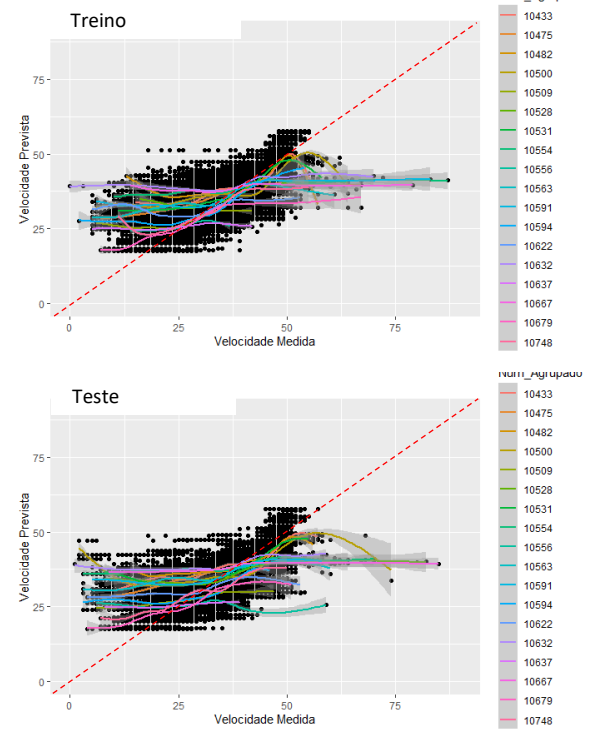
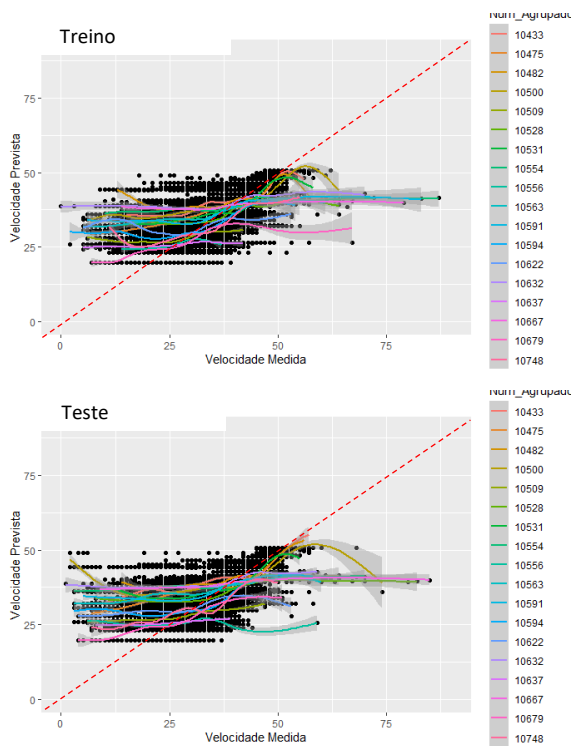
Groups	Name	Variance	Std.Dev.	Corr
Num_Agrupado	(Intercept)	99.3948	9.9697	
	dia_sem1	0.9353	0.9671	-0.33
	dia_sem2	2.2052	1.4850	-0.52 0.72
	dia_sem3	1.6569	1.2872	-0.62 0.81 0.85
	dia_sem4	2.2306	1.4935	-0.64 0.77 0.74 0.88
	dia_sem5	8.4143	2.9007	0.65 -0.02 -0.33 -0.41 -0.40
	dia_sem6	23.2440	4.8212	0.73 -0.31 -0.62 -0.63 -0.60 0.88
	PeriodoP2	1.0378	1.0187	-0.05 -0.37 -0.40 -0.36 -0.31 -0.02 0.28
	PeriodoP3	58.5643	7.6527	-0.77 0.39 0.61 0.68 0.53 -0.67 -0.83
	PeriodoP4	69.3794	8.3294	-0.78 0.41 0.69 0.70 0.65 -0.70 -0.93
	PeriodoP5	77.9679	8.8299	-0.78 0.42 0.61 0.66 0.76 -0.68 -0.89
	PeriodoP6	13.9757	3.7384	-0.60 0.39 0.59 0.52 0.56 -0.53 -0.80
	Residual	21.9472	4.6848	

Number of obs: 121245, groups: Num_Agrupado, 18

Dispersion estimate for gaussian family (sigma^2): 21.9

Conditional model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	15.579	1.849	8.426	<2e-16 ***



Obs. Nas abscissas estão os valores reais e nas ordenadas os valores estimados.

Fonte: o autor

5.2.2.3 Modelos longitudinais

Seguindo com as análises, foram elaborados modelos Longitudinais de Dados em Painel, onde são considerados apenas os efeitos fixos. Uma vez que o objetivo do modelo é o de efetuar estimativas do passo seguinte, foram utilizadas apenas variáveis explicativas gerais – como precipitação de chuva, presença de faixa de ônibus e número de faixas na seção – e variáveis quantitativas com defasagem de 1 a 8 passos.

O uso de variáveis defasadas permite obter o valor da variável dependente no passo seguinte ao e medição, ou seja, estando no instante $t=0$ deve ser possível estimar o valor da variável avaliada no instante $t+1$, com base nas variáveis de $t=0$ à $t-j$, onde j indica a quantidade de defasagens temporais adotadas.

Os modelos longitudinais testados estão apresentados na Tabela 5-10. Como já indicado no item 4.4.3., as variáveis de trecho, que consideram o tempo ou velocidade de percurso incluem muitos dados vazios. De modo similar, as variáveis relativas aos sensores vizinhos também incluem alguns dados vazios. Assim, os modelos que consideraram sensores vizinhos foram elaborados com 118.740 registros e aqueles que consideram o tempo ou velocidade de percurso nos trechos vizinhos foram criados com base em 90.161 registros. Para considerar essa redução, os modelos na base de treino podem ser comparados pela coluna que indica o quociente entre o valor do AIC e a quantidade de registros (observações).

Os modelos longitudinais foram montados de forma incremental, sendo que apenas os dois primeiros foram testados com estrutura de covariância autorregressiva. Princípios pelo modelo L1 que incluiu apenas variáveis gerais: de dia da semana, período do dia, número de faixas de rolamento, existência de faixa exclusiva de ônibus (excluída por não ser estatisticamente significativa), precipitação de chuva e classe da via. Ao modelo L1 acrescentamos as variáveis quantitativas defasadas no próprio sensor, gerando no modelo L2. O modelo L3 acresceu variáveis multiplicativas defasadas relativas ao próprio sensor: velocidade * volume, velocidade * log da ocupação e volume * log da ocupação.

No modelo L4 foram incluídas as variáveis quantitativas defasadas dos sensores vizinhos a montante e jusante, sendo este o primeiro com número menor de registros. No modelo L5 retomamos o modelo L3 e acrescentamos as variáveis quantitativas defasadas dos trechos vizinhos. No modelo L6, também baseado no modelo L3, efetuamos a inclusão de variáveis que foram obtidas como sendo o quociente da divisão da velocidade e do volume nos sensores

vizinhos pela distância destes até o sensor avaliado, incorporando assim a característica espacial do trecho.

Tabela 5-10 – Desempenho dos modelos longitudinais

Modelo	Descrição dos modelo	AIC	Nº de observ.	AIC / observ	Sigma2	MAPE	RMSE	Tempo de treino (s)	Tempo de teste (s)
L1	dia_sem + Período + CH_1h + Classe + Num_faixa	817720,3	121245	6,74	49,9	22,91%	7,513	3,28	3,11
L1_AR	L1 - ARN	585770,1	121245	4,83	7,2	25,05%	8,028	147,85	1904,20
L2	L1 + valores defasados no sensor	583478,8	121245	4,81	7,2	6,62%	2,789	27,24	24,66
L2_AR	L2 - ARN	583398,5	121245	4,81	7,1	6,62%	2,790	79,37	1323,23
L3	L2 + variáveis multiplicativas no sensor	579973,7	121245	4,78	7,0	6,58%	2,750	165,34	159,06
L4	L3 + variáveis defasadas nos sensores vizinhos	567468,5	118740	4,78	6,9	6,46%	2,729	600,68	441,58
L5	L3 + variáveis defasadas nos trechos vizinhos	429218,1	90161	4,76	6,8	6,38%	2,657	184,27	100,64
L6	L4 + variáveis defasadas nos sensores vizinhos / distância	567210,6	118740	4,78	6,9	6,46%	2,715	471,85	954,50
L7	L5 + variáveis defasadas do inverso do tempo de percurso	não convergiu							
L8	L4 + variáveis defasadas do inverso do volume nos sensores vizinhos	não convergiu							

Fonte: o autor

Por fim, nos modelos L7 e L8 buscamos a incorporação de variáveis similares às utilizadas nos modelos STARIMA, qual sejam, variáveis que consideram o inverso do tempo e da velocidade de percurso nos trechos a montante e jusante (L7) e o inverso do volume nos sensores vizinhos (L8). Nestes dois casos não ocorreu convergência do modelo.

Em todos os modelos longitudinais efetuamos a inserção de todas as variáveis de cada tipo e posteriormente efetuamos um processo parcial de *step-up*, ou seja, eliminando parte das variáveis que não apresentassem significância estatística. No caso das variáveis defasadas, optamos por manter as variáveis de intervalos intermediários, ainda que não indicassem relevância estatística no modelo de treino.

Novamente, os modelos autorregressivos apresentaram tempo excessivo de processamento no banco de testes e, como apresentado na Figura 5-17, devido não

apresentarem ganho significativo de acurácia em relação aos modelos não estruturados no banco de testes.

Figura 5-17 – Modelo Longitudinal de Dados em Painel

L2

```

Formula:
vel_media ~ dia_u + Período + CH_1h + Classe + Num_faixa +
vel_media_lag1 +
  vel_media_lag2 + vel_media_lag3 + vel_media_lag4 + vel_media_lag5 +
  vel_media_lag6 + vel_media_lag7 + vel_media_lag8 + volume_lag1 +
  volume_lag2 + volume_lag3 + volume_lag4 + volume_lag5 +
log_ocup_media_lag1 +
  log_ocup_media_lag2 + log_ocup_media_lag3 + log_ocup_media_lag4 +
  log_ocup_media_lag5 + log_ocup_media_lag6 + offset(log(data3))
Data: RD2nna48
      AIC      BIC    logLik deviance df.resid
583478.8 583789.4 -291707.4 583414.8   121244
  
```

Dispersion estimate for gaussian family (sigma^2): 7.19

Conditional model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.360e+01	2.887e-01	-47.11	< 2e-16 ***
dia_u1	1.055e-01	1.713e-02	6.16	7.38e-10 ***
PeríodoP2	-3.726e-02	3.303e-02	-1.13	0.25937
PeríodoP3	-2.732e-01	3.182e-02	-8.59	< 2e-16 ***
PeríodoP4	-3.531e-02	3.184e-02	-1.11	0.26753
PeríodoP5	-4.144e-02	3.235e-02	-1.28	0.20017
PeríodoP6	2.827e-01	3.162e-02	8.94	< 2e-16 ***
CH_1h	-1.383e-05	5.978e-06	-2.31	0.02069 *
ClasseExpressa	5.033e-01	3.110e-02	16.18	< 2e-16 ***
Num_faixa4	2.058e-03	1.790e-02	0.12	0.90843
Num_faixa5	2.919e-01	4.222e-02	6.91	4.71e-12 ***
Num_faixa6	-5.533e-01	5.603e-02	-9.88	< 2e-16 ***
vel_media_lag1	4.694e-01	3.748e-03	125.26	< 2e-16 ***
vel_media_lag2	1.750e-01	3.981e-03	43.95	< 2e-16 ***
vel_media_lag3	9.534e-02	4.027e-03	23.67	< 2e-16 ***
vel_media_lag4	3.420e-02	4.021e-03	8.51	< 2e-16 ***
vel_media_lag5	9.540e-02	4.009e-03	23.80	< 2e-16 ***
vel_media_lag6	-1.567e-02	4.002e-03	-3.92	9.03e-05 ***
vel_media_lag7	4.388e-02	3.207e-03	13.68	< 2e-16 ***
vel_media_lag8	5.335e-02	2.844e-03	18.76	< 2e-16 ***
volume_lag1	8.689e-04	3.729e-04	2.33	0.01979 *
volume_lag2	-7.698e-04	4.036e-04	-1.91	0.05658 .
volume_lag3	-6.612e-04	4.115e-04	-1.61	0.10808
volume_lag4	-2.407e-03	4.036e-04	-5.96	2.46e-09 ***
volume_lag5	1.123e-03	3.726e-04	3.01	0.00258 *
log_ocup_media_lag1	-9.481e-01	4.745e-02	-19.98	< 2e-16 ***
log_ocup_media_lag2	-2.856e-01	4.823e-02	-5.92	3.20e-09 ***
log_ocup_media_lag3	1.921e-01	4.879e-02	3.94	8.25e-05 ***
log_ocup_media_lag4	4.721e-01	4.879e-02	9.68	< 2e-16 ***
log_ocup_media_lag5	2.185e-01	4.811e-02	4.54	5.56e-06 ***
log_ocup_media_lag6	1.955e-01	4.699e-02	4.16	3.19e-05 ***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

L2_AR

```

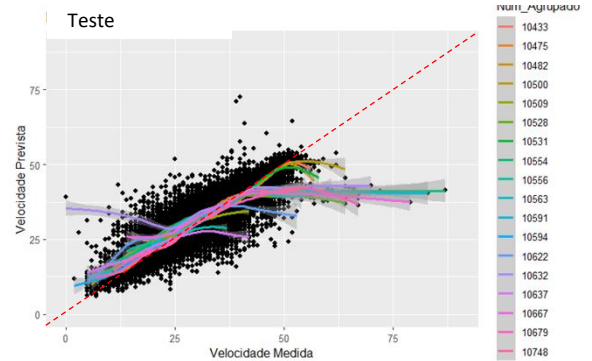
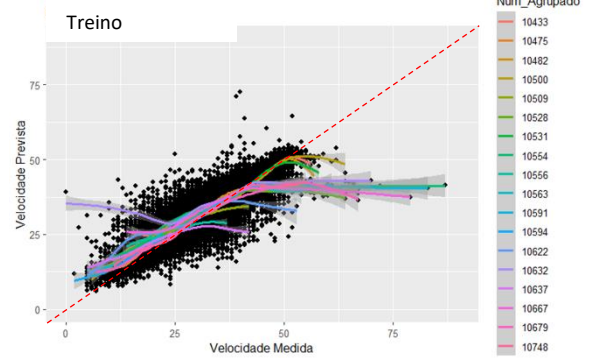
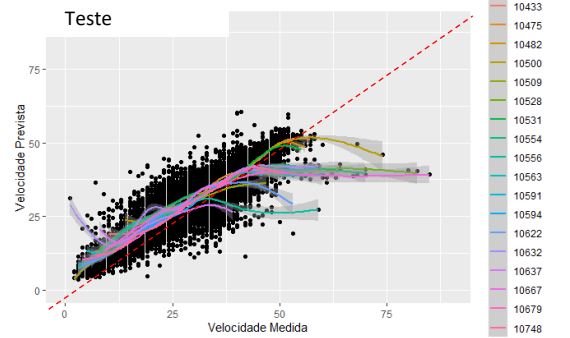
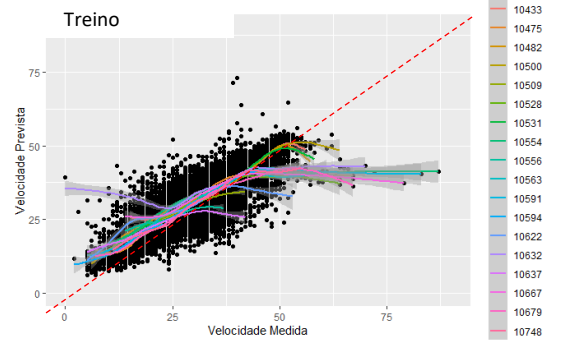
Formula:
Vel_media ~ dia_u + Período + CH_1h + Classe + Num_faixa +
Vel_media_lag1 +
  Vel_media_lag2 + Vel_media_lag3 + Vel_media_lag4 + Vel_media_lag5 +
  Vel_media_lag6 + Vel_media_lag7 + Vel_media_lag8 + volume_lag1 +
  Volume_lag2 + Volume_lag3 + Volume_lag4 + Volume_lag5 + log_ocup_media_lag1 +
  log_ocup_media_lag2 + log_ocup_media_lag3 + log_ocup_media_lag4 +
  log_ocup_media_lag5 + log_ocup_media_lag6 + offset(log(data3))
Data: RD2nna48
      AIC      BIC    logLik deviance df.resid
583478.8 583789.4 -291707.4 583414.8   121244
  
```

Dispersion estimate for gaussian family (sigma^2): 7.19

Conditional model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.360e+01	2.887e-01	-47.11	< 2e-16 ***
dia_u1	1.055e-01	1.713e-02	6.16	7.38e-10 ***
PeríodoP2	-3.726e-02	3.303e-02	-1.13	0.25937
PeríodoP3	-2.732e-01	3.182e-02	-8.59	< 2e-16 ***
PeríodoP4	-3.531e-02	3.184e-02	-1.11	0.26753
PeríodoP5	-4.144e-02	3.235e-02	-1.28	0.20017
PeríodoP6	2.827e-01	3.162e-02	8.94	< 2e-16 ***
CH_1h	-1.383e-05	5.978e-06	-2.31	0.02069 *
ClasseExpressa	5.033e-01	3.110e-02	16.18	< 2e-16 ***
Num_faixa4	2.058e-03	1.790e-02	0.12	0.90843
Num_faixa5	2.919e-01	4.222e-02	6.91	4.71e-12 ***
Num_faixa6	-5.533e-01	5.603e-02	-9.88	< 2e-16 ***
Vel_media_lag1	4.694e-01	3.748e-03	125.26	< 2e-16 ***
Vel_media_lag2	1.750e-01	3.981e-03	43.95	< 2e-16 ***
Vel_media_lag3	9.534e-02	4.027e-03	23.67	< 2e-16 ***
Vel_media_lag4	3.420e-02	4.021e-03	8.51	< 2e-16 ***
Vel_media_lag5	9.540e-02	4.009e-03	23.80	< 2e-16 ***
Vel_media_lag6	-1.567e-02	4.002e-03	-3.92	9.03e-05 ***
Vel_media_lag7	4.388e-02	3.207e-03	13.68	< 2e-16 ***
Vel_media_lag8	5.335e-02	2.844e-03	18.76	< 2e-16 ***
Volume_lag1	8.689e-04	3.729e-04	2.33	0.01979 *
Volume_lag2	-7.698e-04	4.036e-04	-1.91	0.05658 .
Volume_lag3	-6.612e-04	4.115e-04	-1.61	0.10808
Volume_lag4	-2.407e-03	4.036e-04	-5.96	2.46e-09 ***
Volume_lag5	1.123e-03	3.726e-04	3.01	0.00258 *
log_ocup_media_lag1	-9.481e-01	4.745e-02	-19.98	< 2e-16 ***
log_ocup_media_lag2	-2.856e-01	4.823e-02	-5.92	3.20e-09 ***
log_ocup_media_lag3	1.921e-01	4.879e-02	3.94	8.25e-05 ***
log_ocup_media_lag4	4.721e-01	4.879e-02	9.68	< 2e-16 ***
log_ocup_media_lag5	2.185e-01	4.811e-02	4.54	5.56e-06 ***
log_ocup_media_lag6	1.955e-01	4.699e-02	4.16	3.19e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Obs. Nas abscissas estão os valores reais e nas ordenadas os valores estimados

Fonte: o autor

Na Figura 5-17 constata-se que as linhas coloridas, representando a média da previsão para cada sensor, são consideravelmente mais próximas à diagonal do gráfico do que aquelas apresentadas na Figura 5-16. O aumento de acurácia confirma os valores dos medidores de acurácia da Tabela 5-10.

Com relação às variáveis utilizadas, verificou-se que as variáveis gerais utilizadas no modelo L1 se mantiveram nos demais modelos. Quanto aos modelos com a variável “ocupancia” (soma das ocupações durante o intervalo), foi necessário efetuar a transformação para o formato logarítmico.

A inclusão da variável relacionada à chuva se mostrou significativa, sendo que em todos os casos a ocorrência de chuva implicava na redução da velocidade média, como de fato se verifica no mundo real.

O uso de variáveis multiplicativas, das características locais e daquelas dos sensores vizinhos pelo inverso da distância entre sensores, mostrou-se adequado, propiciando modelos com bons resultados de acurácia.

A estratégia de se utilizar dados defasados resultou em variáveis significativas e que aprimoraram a acurácia das estimações, permitindo incluir, de modo mais explícito, o reflexo da situação do tráfego nos intervalos anteriores.

Dentre os modelos Longitudinais, sem covariância por autorregressão, destacaram-se aqueles que consideraram os sensores e trechos vizinhos. Observa-se, no entanto, que o modelo L3, que considerou apenas os dados do próprio sensor, apresentou desempenho bastante razoável e próximo aos modelos que consideraram os vizinhos.

5.2.2.4 Modelos hierárquicos completos

Por fim, foram elaborados quatro modelos HLM, que incluíram: o aninhamento por sensor e variáveis de efeito fixo, sendo dois deles com a variável período nos efeitos aleatórios.

Considerando a semelhança na acurácia obtida entre os modelos longitudinais L3, L4, L5 e L6 (Tabela 5-10), optamos por evoluir o modelo L3, que considera apenas os dados do próprio sensor, e o L6, que considera características dos sensores vizinhos e a distância dos trechos. Estes são mais fáceis de implementar num modelo de produção real do que os

modelos que consideram tempo ou velocidade de percurso. Isso se deve à necessidade de elaboração de série temporal específica para os trechos, grande quantidade de valores ausentes nesse tipo de variável. A

Tabela 5-11 apresenta os 4 modelos testados, comparando-os com os modelos longitudinais correspondentes.

Tabela 5-11 – Desempenho dos modelos longitudinais

Modelo	Descrição dos modelo	AIC	Nº de observ.	AIC / observ.	Sigma2	MAPE	RMSE	Tempo de treino (s)	Tempo de teste (s)
L3	Modelo Longitudinal L3	579973,7	121245	4,78	7,0	6,58%	2,750	165,34	159,06
L3_N1N	L3 aninhado por sensor	579074,3	121245	4,78	6,9	6,65%	2,734	164,78	126,33
L3_N1N_A2	L3 aninhado por sensor e com variável aleatória Período	577845,9	121245	4,77	6,8	6,72%	2,739	355,11	93,44
L6	Modelo Longitudinal L6	567210,6	118740	4,78	6,9	6,46%	2,715	471,85	344,39
L6_N1N	L6 aninhado por sensor	566356,2	118740	4,77	6,8	6,52%	2,708	650,40	349,05
L6_N1N_A2	L6 aninhado por sensor e com variável aleatória Período	564976,2	118740	4,76	6,7	6,56%	2,700	797,08	380,26

Observa-se que os modelos com base no modelo L6 apresentaram tempo de processamento maior, todavia, considerado viável para a aplicação em tempo real, já que foram gastos cerca de seis minutos para o processamento de 11 dias de dados. Pela coluna comparativa (AIC/observações) verifica-se que todos os modelos obtiveram acurácia semelhante na base de treino. A comparação pelos medidores de acurácia MAPE e RMSE indica certa vantagem para os modelos que consideram os vizinhos (base L6).

Destaca-se que os modelos hierárquicos com variáveis aleatórias incluindo os dias úteis (A1) e os dias da semana (A3) foram testados, todavia não ocorreu convergência. Assim, adotamos o uso do modelo A2 que considera apenas a variável Período nos efeitos aleatórios.

Tendo em vista a similaridade dos diversos modelos, na Figura 5-18, apresentamos apenas o modelo L6_N1N_A2, que é o mais completo. Neste caso incluímos também dois gráficos suplementares que apresentam as previsões para cada sensor e para cada eixo de modo separado, sendo possível verificar a adequação do modelo de modo mais detalhado.

Figura 5-18 – Modelos HLM e Efeitos Aleatórios de Tempo Sem e Com Autocorrelação

L6_N1N_A2

Formula: $Vel_media \sim dia_u1 + CH_1h + Classe + Num_faixa + Vel_media_lag1 +$
 $Vel_media_lag2 + Vel_media_lag3 + Vel_media_lag4 + Vel_media_lag5 +$
 $Vel_media_lag6 + Vel_media_lag7 + Vel_media_lag8 + Volume_lag1 +$
 $Volume_lag2 + Volume_lag3 + Volume_lag4 + log_ocup_media_lag1 +$
 $log_ocup_media_lag2 + log_ocup_media_lag3 + log_ocup_media_lag4 +$
 $log_ocup_media_lag5 + log_ocup_media_lag6 + oclog_vol_lag1 +$
 $oclog_vol_lag2 + oclog_vol_lag3 + oclog_vol_lag4 + oclog_vol_lag5 +$
 $oclog_vol_lag6 + oclog_vol_lag7 + oclog_vol_lag8 + vel_oclog_lag1 +$
 $vel_oclog_lag2 + vel_oclog_lag3 + vel_oclog_lag4 + vel_oclog_lag5 +$
 $vel_oclog_lag6 + vel_oclog_lag7 + vel_vol_lag1 + vel_vol_lag2 +$
 $vel_vol_lag3 + vel_vol_lag4 + vel_vol_lag5 + vel_vol_lag6 +$
 $vel_vol_lag7 + vel_vol_lag8 + Ocup_md_RM_lag1 + Ocup_md_RM_lag2 +$
 $Ocup_md_RM_lag3 + Ocup_md_RM_lag4 + Ocup_md_RM_lag5 + Ocup_md_RM_lag6 +$
 $Ocup_md_RM_lag7 + Ocup_md_RM_lag8 + Vel_RJ_Dist_lag1 + Vel_RJ_Dist_lag2 +$
 $Vel_RJ_Dist_lag3 + Vel_RJ_Dist_lag4 + Vel_RJ_Dist_lag5 +$
 $Vel_RJ_Dist_lag6 + Vel_RJ_Dist_lag7 + Vel_RJ_Dist_lag8 +$
 $Vel_RM_Dist_lag1 + Vel_RM_Dist_lag2 + Vel_RM_Dist_lag3 +$
 $Vel_RM_Dist_lag4 + Vel_RM_Dist_lag5 + Vel_RM_Dist_lag6 +$
 $Vel_RM_Dist_lag7 + Vel_RM_Dist_lag8 + (Período | Num_Agrupado) +$
 $offset(log(data3))$
 Data: RD2na48

	AIC	BIC	logLik	deviance	df.resid
	564976.2	565964.1	-282386.1	564772.2	118718

Random effects:

Conditional model:

Groups	Name	Variance	Std.Dev.	Corr
Num_Agrupado	(Intercept)	1.65104	1.2849	
	PeriodoP2	0.05775	0.2403	-0.65
	PeriodoP3	0.79744	0.8930	-0.85 0.52
	PeriodoP4	0.39268	0.6266	-0.77 0.54 0.77
	PeriodoP5	0.41407	0.6435	-0.77 0.66 0.70 0.94
	PeriodoP6	0.10848	0.3294	-0.14 0.46 0.27 0.58 0.52
Residual		6.74350	2.5968	

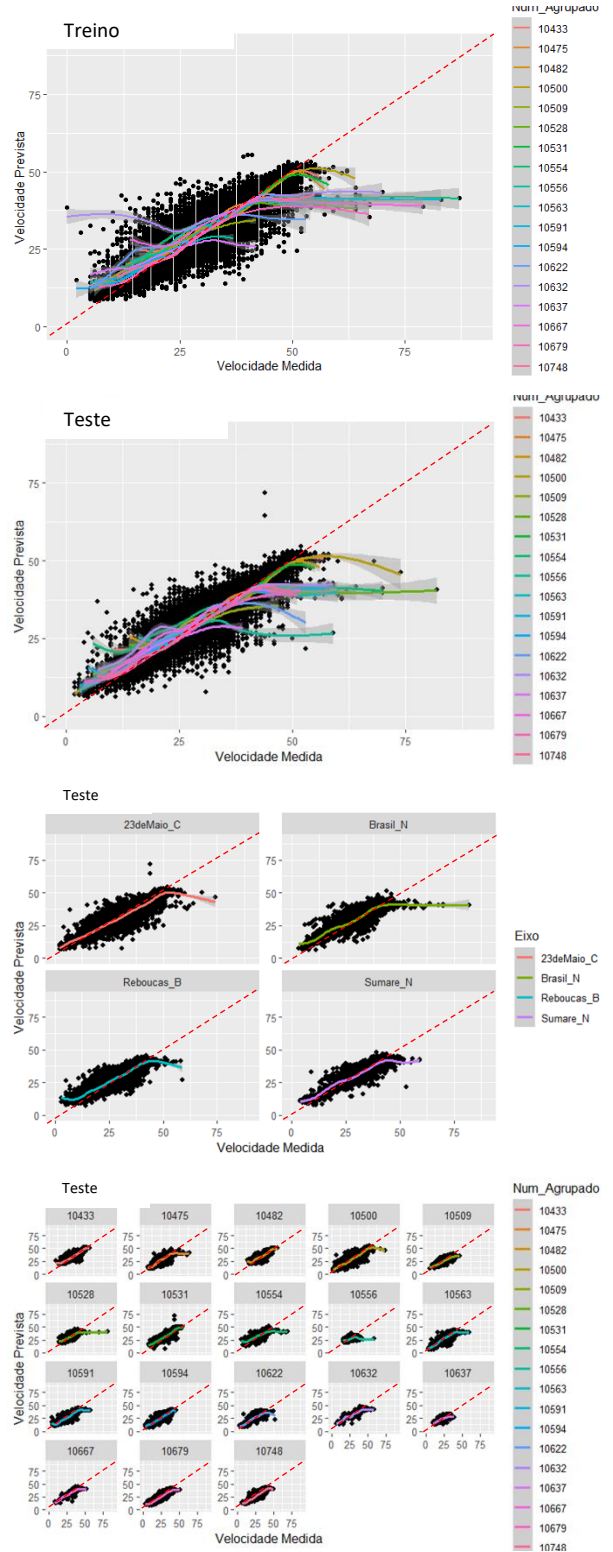
Number of obs: 118740, groups: Num_Agrupado, 18

Dispersion estimate for gaussian family (sigma^2): 6.74

Conditional model:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.818e+00	7.799e-01	-3.613	0.000303 ***
dia_u1	1.482e-01	1.823e-02	8.128	4.37e-16 ***
CH_1h	-3.841e-05	6.575e-06	-5.842	5.16e-09 ***
ClasseExpressa	-1.899e-01	8.070e-01	-0.235	0.814000
Num_faixa4	2.336e-01	3.870e-01	0.604	0.545991
Num_faixa5	6.224e-01	7.919e-01	0.786	0.431831
Num_faixa6	-1.919e+00	5.926e-01	-3.239	0.001201 **
Vel_media_lag1	1.988e-01	6.522e-03	30.477	< 2e-16 ***
Vel_media_lag2	9.406e-02	6.571e-03	14.315	< 2e-16 ***
Vel_media_lag3	9.192e-02	6.602e-03	13.922	< 2e-16 ***
Vel_media_lag4	6.347e-02	6.583e-03	9.642	< 2e-16 ***
Vel_media_lag5	8.672e-02	6.561e-03	13.218	< 2e-16 ***
Vel_media_lag6	4.167e-02	6.529e-03	6.383	1.74e-10 ***
Vel_media_lag7	5.350e-02	5.225e-03	10.239	< 2e-16 ***
Vel_media_lag8	5.466e-02	4.005e-03	13.646	< 2e-16 ***
Volume_lag1	4.043e-03	4.095e-03	0.987	0.323435
Volume_lag2	1.024e-02	4.230e-03	2.410	0.015946 *
Volume_lag3	1.646e-03	4.241e-03	0.388	0.697873
Volume_lag4	-1.125e-02	4.094e-03	-2.748	0.005997 **
log_ocup_media_lag1	-1.421e+00	6.128e-02	-23.194	< 2e-16 ***
log_ocup_media_lag2	-6.393e-01	6.266e-02	-10.203	< 2e-16 ***
log_ocup_media_lag3	1.134e-01	6.297e-02	1.802	0.071619 *
log_ocup_media_lag4	4.801e-01	6.274e-02	7.652	1.98e-14 ***
log_ocup_media_lag5	4.436e-02	5.436e-02	0.816	0.414493
log_ocup_media_lag6	4.206e-01	5.309e-02	7.922	2.34e-15 ***
oclog_vol_lag1	-5.545e-03	5.096e-04	-10.881	< 2e-16 ***
oclog_vol_lag2	-5.509e-04	5.244e-04	-1.051	0.293459
oclog_vol_lag3	4.682e-04	5.225e-04	0.896	0.370187
oclog_vol_lag4	1.311e-04	5.074e-04	2.583	0.009189 **
oclog_vol_lag5	9.775e-04	1.247e-04	7.837	4.62e-15 ***
oclog_vol_lag6	1.539e-04	1.237e-04	1.244	0.213532
oclog_vol_lag7	2.364e-04	1.214e-04	1.947	0.051495 *
oclog_vol_lag8	1.343e-04	1.093e-04	1.229	0.219187
vel_oclog_lag1	1.781e-02	8.678e-04	20.522	< 2e-16 ***
vel_oclog_lag2	8.649e-03	8.966e-04	10.162	< 2e-16 ***
vel_oclog_lag3	5.841e-04	9.046e-04	0.646	0.518444
vel_oclog_lag4	-5.316e-03	9.034e-04	-5.885	3.99e-09 ***
vel_oclog_lag5	3.204e-03	8.646e-04	3.706	0.000211 ***
vel_oclog_lag6	-6.471e-03	8.530e-04	-7.587	3.28e-14 ***
vel_oclog_lag7	-8.285e-04	7.690e-04	-1.077	0.281290
vel_vol_lag1	8.806e-04	2.821e-05	31.220	< 2e-16 ***
vel_vol_lag2	-2.294e-04	3.102e-05	-7.393	1.43e-13 ***
vel_vol_lag3	-1.680e-04	3.116e-05	-5.394	6.90e-08 ***
vel_vol_lag4	2.735e-05	3.044e-05	0.899	0.368861
vel_vol_lag5	-1.512e-04	2.044e-05	-7.394	1.43e-13 ***
vel_vol_lag6	-2.172e-05	2.024e-05	-1.073	0.283167
vel_vol_lag7	-6.051e-05	1.972e-05	-3.068	0.002154 **
vel_vol_lag8	-2.074e-05	1.705e-05	-1.212	0.223741
Ocup_md_RM_lag1	5.765e-07	2.327e-06	0.248	0.804311
Ocup_md_RM_lag2	2.041e-06	2.327e-06	0.877	0.380287
Ocup_md_RM_lag3	7.644e-06	2.327e-06	3.285	0.001020 **
Ocup_md_RM_lag4	-2.544e-06	2.327e-06	-1.094	0.274138
Ocup_md_RM_lag5	8.300e-06	2.326e-06	3.569	0.000358 ***
Ocup_md_RM_lag6	-8.781e-06	2.324e-06	-3.778	0.000158 ***
Ocup_md_RM_lag7	5.616e-06	2.321e-06	2.419	0.015562 **
Ocup_md_RM_lag8	9.301e-06	2.321e-06	4.007	6.14e-05 ***
Vel_RJ_Dist_lag1	-1.034e-01	1.844e-02	-5.609	2.04e-08 ***
Vel_RJ_Dist_lag2	-2.956e-02	1.573e-02	-1.879	0.060270 *
Vel_RJ_Dist_lag3	1.140e-01	2.088e-02	5.459	4.78e-08 ***
Vel_RJ_Dist_lag4	5.205e-03	2.191e-02	0.238	0.812175
Vel_RJ_Dist_lag5	3.589e-02	1.794e-02	2.001	0.045440 *
Vel_RJ_Dist_lag6	-6.428e-02	1.809e-02	-3.553	0.000381 ***
Vel_RJ_Dist_lag7	4.842e-02	2.160e-02	2.241	0.025006 *
Vel_RJ_Dist_lag8	-4.566e-02	1.996e-02	-2.288	0.022163 **
Vel_RM_Dist_lag1	-1.861e-01	2.553e-02	-7.290	3.09e-13 ***
Vel_RM_Dist_lag2	-5.419e-02	2.712e-02	-1.998	0.045709 *
Vel_RM_Dist_lag3	1.591e-01	2.780e-02	5.722	1.05e-08 ***
Vel_RM_Dist_lag4	-1.395e-02	2.759e-02	-0.506	0.613179
Vel_RM_Dist_lag5	3.483e-01	2.732e-02	12.749	< 2e-16 ***
Vel_RM_Dist_lag6	-2.123e-01	2.756e-02	-7.706	1.30e-14 ***
Vel_RM_Dist_lag7	2.248e-02	2.737e-02	0.821	0.411484
Vel_RM_Dist_lag8	-1.327e-02	2.526e-02	-0.525	0.599439
Vol_RM_Dist_lag1	-5.192e-02	3.740e-03	-13.884	< 2e-16 ***
Vol_RM_Dist_lag2	-1.397e-02	3.811e-03	-3.666	0.000247 ***
Vol_RM_Dist_lag3	2.080e-02	3.880e-03	5.360	8.30e-08 ***
Vol_RM_Dist_lag4	8.676e-04	3.835e-03	0.226	0.821016
Vol_RM_Dist_lag5	-1.501e-03	3.836e-03	-0.391	0.695508
Vol_RM_Dist_lag6	2.864e-02	3.868e-03	7.406	1.30e-13 ***
Vol_RM_Dist_lag7	-1.257e-03	3.786e-03	-0.333	0.737817
Vol_RM_Dist_lag8	1.012e-02	3.727e-03	2.716	0.006605 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Obs. Nas abscissas estão os valores reais e nas ordenadas os valores estimados

Fonte: o autor

5.2.3 Comparando os modelos de estimação da velocidade

Quanto à aplicação dos modelos avaliados, verificamos que as estimações com o modelo STARIMA implicaram na necessidade de montagem da matriz espacial, antes da elaboração do modelo, sendo esta dependente da escolha de parâmetros das ordens espacial e temporal feita através da análise de gráficos de autocorrelação dos dados. Já o modelo de Dados em Painel permite aplicação mais direta, contudo, é necessária a definição das variáveis a serem utilizadas e dos níveis de aninhamento a serem considerados, o que pode demandar um grande esforço de pesquisa.

A fim de comparar os tipos de estimação, elaboramos a Tabela 5-12, na qual incluímos 2 tipos de estimação baseados no modelo STARIMA e outros 2 baseados em Dados em Painel. A comparação se dá apenas para a previsão da velocidade pontual, pois apenas essa característica foi tratada através de Dados em Painel. Foram escolhidos os dias 20 e 21/03/2018 que eram dias presentes nos testes dos dois modelos. O horizonte de previsão é de 1 passo, que é o principal objetivo do modelo de estimação.

Como o dia 20/03 foi um dia atípico, com chuvas muito fortes e grande impacto no tráfego, a Tabela 5-12 apresenta a acurácia durante o período de impacto da chuva (das 15h às 21h do dia 20/03) separadamente do restante do período de testes, além de indicar a acurácia geral incluindo os dois dias completos.

Tabela 5-12 – Comparação da acurácia dos modelos de estimação de velocidade para os ambientes de teste

Medidor de Acurácia	Situação	Painel L6	Painel L6_N1N_A2	STARIMA 1-1/T	STARIMA 1/D
MAPE	Chuvas Fortes (Atípico)	0,1663	0,1889	0,1558	0,1569
	Exceto Chuvas Fortes (Típico)	0,0666	0,0657	0,0696	0,0696
	Geral	0,0786	0,0811	0,0804	0,805
RMSE	Chuvas Fortes (Atípico)	3,724	3,818	3,708	3,720
	Exceto Chuvas Fortes (Típico)	2,841	2,806	3,022	3,022
	Geral	2,966	2,951	3,116	3,118

Considera apenas os dias 20 e 21/03/2018

Fonte: o autor

Da Tabela 5-12 é possível verificar que as estimações com STARIMA desempenharam melhor no período com chuvas fortes. Importante notar a grande variação de acurácia apresentada pelo medidor MAPE, em relação aos outros períodos, e que não aparece de modo tão acentuado no medidor RMSE. Isto está relacionado à forma de medição utilizada. O MAPE mede a diferença percentual entre o medido e o previsto, assim, para valores muito baixos, como os que ocorreram nesse período, a variação percentual é alta. Por exemplo, para um valor medido de 5 km/h, como ocorreu no nosso caso, uma previsão de 7km/h (um erro de apenas 2km/h a mais) representa um índice de 40%. Já o RMSE indica o erro de modo quantitativo em relação à diferença entre o valor medido e o valor previsto. Entendemos que, neste caso, a comparação pelo RMSE é mais adequada, já que o impacto das chuvas causa grande variação nos dados medidos.

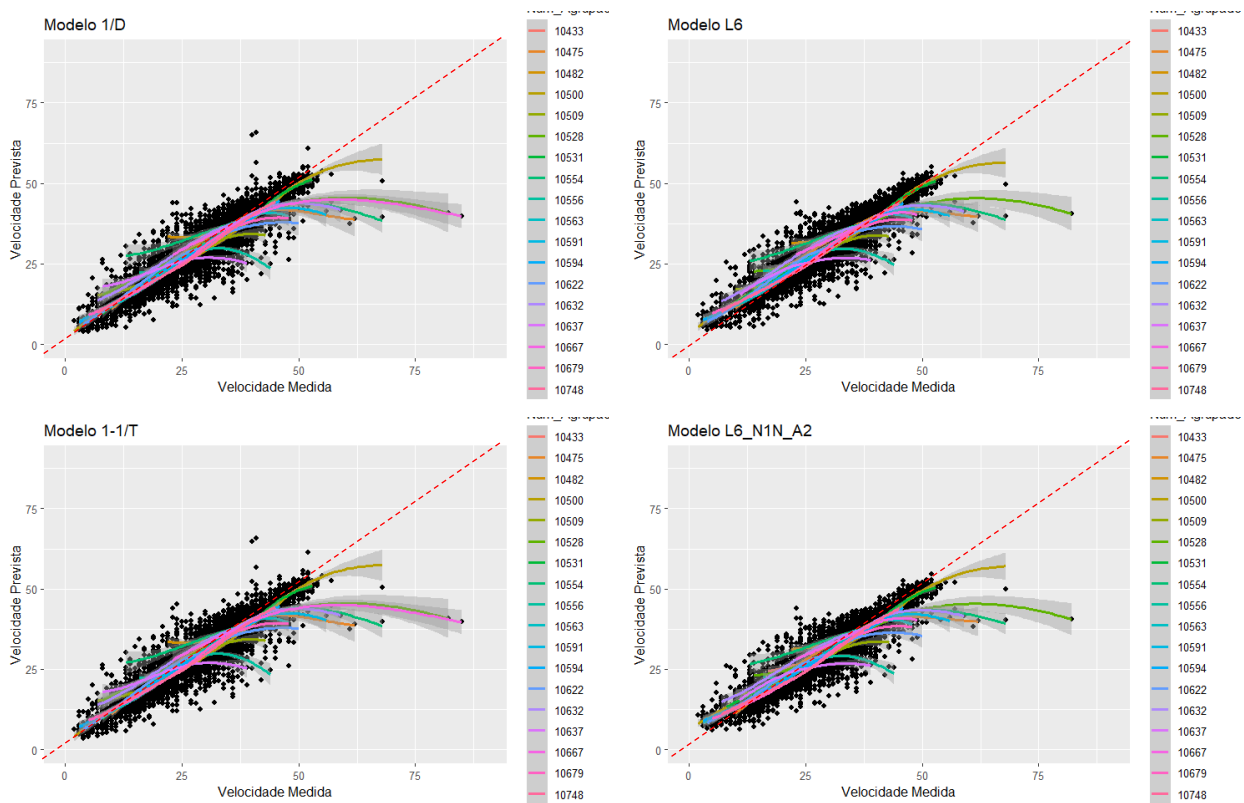
Para o período sem chuvas, as estimações com os modelos de Dados em Painel apresentaram-se mais acuradas, tanto na comparação pelo MAPE como pelo RMSE. Quando consideradas as 48 horas de modo geral, verificamos que a situação se inverte e os modelos com Dados em Painel se apresentam mais acurados nos dois medidores de acurácia.

Dentre os dois tipos de estimacão com STARIMA verifica-se que o tipo 1-1/T, cuja matriz espacial é assimétrica e considera as condições do fluxo, apresentou acurácia ligeiramente melhor. No caso dos modelos com Dados em Painel, o modelo Longitudinal (L6) apresentou melhor acurácia no período de chuvas e o modelo hierárquico com variáveis fixas e aleatórias (L6_N1N_A2) apresentou-se melhor no restante do período e no período de 48 horas.

Para melhor visualização dos resultados obtidos, com cada um dos tipos de estimacão, foram elaborados os gráficos da Figura 5-19, onde é possível verificar que as estimacões pelos diversos modelos apresentam dispersão similar, sendo os modelos de Dados em Painel ligeiramente menos dispersos.

Assim, é possível concluir que, no caso da característica “velocidade pontual”, tanto o STARIMA como os Dados em Painel se mostraram viáveis para aplicaões de estimacão.

Figura 5-19 – Comparação dos Tipos de Estimação no ambiente de teste



Obs. Velocidades medidas em dm/s. Nas abscissas estão os valores reais e nas ordenadas os valores estimados
 Fonte: o autor

5.3 CONSIDERAÇÕES FINAIS DO CAPÍTULO 5

Neste capítulo apresentamos os estudos efetuados, incluindo a aplicação dos métodos de estimação, os modelos elaborados e testados e os resultados obtidos. A forma de aplicação dos métodos foi apresentada de modo detalhado, possibilitando sua aplicação em estudos futuros.

Com relação às estimções de volume de tráfego, baseadas no modelo STARIMA, ficou demonstrada a possibilidade prática da utilização do modelo de matriz agrupada com resultados bastante adequados ao uso. Como indicado anteriormente, a aplicação desse tipo de matriz é inédita nas aplicações de estimação e previsão de tráfego com o modelo STARIMA. Nessa mesma direção, a superioridade das matrizes assimétricas, verificada durante os estudos de matrizes, se confirmou quando dos testes comparativos.

A criação dos modelos com Dados em Painel foi efetuada de modo sequenciado, permitindo o entendimento de cada parcela do modelo. A análise dos diversos modelos

envolve a execução de testes, incluindo e retirando variáveis, o que pode implicar num grande esforço de pesquisa.

A variável de índice pluviométrico na última hora (CH_1), quando inserida nos modelos de dados em painel, apresentou significância estatística. O mesmo ocorreu com a classe da via. As características da via, quantidade de faixas e existência de faixa de ônibus, se mostrou mais significativa em modelos mais simples. Todavia, em modelos com mais variáveis defasadas estas características foram perdendo a significância.

As variáveis relativas ao tempo e à velocidade de percurso, apresentaram à ausência de informações em mais de 25% dos intervalos utilizados. Para os modelos com Dados em Painel isso implicou na redução dos bancos de dados utilizados na pesquisa. Dessa forma, embora seja útil à acurácia dos modelos, sua aplicação deve ser avaliada com cuidado e a possibilidade de sua substituição pela variável de distância deve ser considerada. Isso foi feito na pesquisa através da comparação dos modelos L5 e L6.

As melhores estimações de velocidade, feitas com o uso dos dois modelos avaliados, STARIMA e Dados em Painel, apresentaram acurácia semelhante, com alguma superioridade para os Dados em Painel, cuja estrutura é multivariada. Dessa forma, os estudos realizados indicam a adequação dos dois modelos escolhidos para a pesquisa, com relação à estimação do tráfego.

Os resultados obtidos com o Painel Longitudinal, sem aninhamento, surpreenderam quanto à acurácia, similar à das estimações com aninhamento, sendo superiores no período atípico do teste comparativo. Nessa mesma direção, a inclusão dos efeitos fixos nos tipos de estimação com aninhamento, propiciou melhora em sua acurácia, o que confirma a necessidade de inclusão das variáveis neste tipo de efeito.

Todavia, devemos considerar que os modelos com STARIMA se comportaram melhor no caso de situações atípicas, com mudanças bruscas na velocidade do tráfego. Existe ainda a possibilidade de melhoria dos cenários obtidos com o modelo STARIMA, através da criação de modelos específicos para cada período do dia, de modo a incluir a sazonalidade decorrente, semelhantemente à inclusão da variável período feita nos modelos com Dados em Painel.

No caso dos modelos com Dados em Painel, a inclusão de mais vias, incorporando maior variedade nas características de infraestrutura, e o aumento do banco de dados de treino, podem potencialmente melhorar os modelos elaborados.

6 CONCLUSÕES

A metodologia proposta permite o processo de estimação de variáveis quantitativas do tráfego, de modo que o objetivo principal da pesquisa foi atingido. Ela é composta de etapas, que vão desde a recepção dos dados até sua disponibilização para uso em outros sistemas. Ao longo da pesquisa foram elaborados scripts em Python e R que permitiram a abertura dos dados dos sensores, a construção das séries temporais e a aplicação dos dois modelos estatísticos.

Os trabalhos desenvolvidos nesta tese demonstraram a possibilidade de efetuar a estimação das variáveis quantitativas do tráfego, a partir de modelos estatísticos consolidados na literatura, como STARIMA e Dados em Painel. Os dois modelos utilizados incluem formas de considerar as relações entre os sensores, incorporando a característica espacial, sendo através do uso da matriz espacial no modelo STARIMA, ou através do aninhamento dos dados no caso do uso do modelo de Dados em Painel. Ambos estão baseados na análise de séries temporais, utilizando dados históricos como forma de construção dos modelos.

Confirmando o que se verificou na bibliografia, o modelo STARIMA se mostrou útil, tendo boa acurácia com poucos parâmetros. Embora o desenvolvimento do modelo seja trabalhoso e envolva tempos longos de processamento, sua aplicação na estimação das variáveis quantitativas do tráfego apresenta tempos bastante reduzidos e adequados ao uso em tempo próximo ao real.

Os modelos com Dados em Painel melhoraram com o aumento na quantidade de variáveis explicativas significativas. Todavia, o tempo de processamento também aumenta, tanto na base de treinos como na base de testes. Os modelos autorregressivos aplicados nos modelos de Dados em Painel apresentaram tempos muito maiores de processamento na etapa de estimação, indicando não serem adequados às aplicações em tempo próximo ao real.

A comparação da acurácia obtida para a Velocidade Pontual, demonstrou vantagem aos cenários baseados em Dados em Painel nos períodos típicos e para os modelos STARIMA nos momentos atípicos com chuva. Pela tabela 5-12 pode-se ver que, nos períodos típicos o RMSE foi de 2,806 com Dados em Painel e 3,022 para o STARIMA e no período atípico foi de 3,818 e 3,720 respectivamente. A comparação com MAPE indica no período típico 6,57% para Dados

em Painel e 6,96% para STARIMA e no período atípico respectivamente 15,58% e 16,63%. Todavia, os testes comparativos consideraram apenas dois dias de dados e, portanto, não devem ser considerados de modo conclusivo. A ampliação da base comparativa pode ser tema de pesquisas futuras.

As pesquisas realizadas com as matrizes de pesos espaciais permitiram o entendimento do processo de sua construção e aplicação em modelos autorregressivos. A abordagem dada no estudo, com o uso de matrizes assimétricas e agrupadas, se mostrou inédita no caso dos estudos de previsão de tráfego com STARIMA. Os resultados de previsões de volume, englobando todos os 18 sensores, indicaram que a matriz agrupada superou as demais em 42% dos testes, sendo que a vizinhança por contiguidade foi a melhor em 33% das vezes (Tabela 5-5). Na mesma direção, reunidas as contribuições das matrizes assimétricas, estas apresentaram-se mais acuradas que as simétricas em 75,4% dos cenários testados (Tabela 5-6).

Durante os estudos de matrizes testou-se a possibilidade de uso do modelo STARIMA para previsões do volume com horizontes de 5 e 15 intervalos à frente (20 e 60 min). Neste caso os testes foram limitados apenas ao período da tarde. Conforme é possível verificar na Tabela 5-3, os melhores resultados medidos através do MAPE, nos dias típicos, para a estimação (1 intervalo) foram de 6,9% na via expressa e 10,7% na arterial, para as previsões de 5 intervalos a acurácia foi de 9,2% e 8,7% e para 15 intervalos foi obtida acurácia de 9,6% e 14,8%. Já no dia atípico, de véspera de feriado, a estimação teve acurácia de 14,1 e 14,4%, as previsões de 5 intervalos obtiveram 16,3% e 14,3% e as de 15 intervalos 22,3% e 16,4%. Assim, verifica-se que, apesar da perda de acurácia, para horizontes de tempo maiores a aplicação do modelo parece ser possível. Novos estudos, com maior quantidade de dados, devem confirmar essa análise.

Os estudos preliminares com Dados em Painel para a variável quantitativa Volume, realizados paralelamente à pesquisa e ainda em andamento, foram apresentados resumidamente no Apêndice B, sendo possível verificar na Figura B-2 as linhas médias para cada sensor e a dispersão de dados que se aproximam da reta diagonal, confirmando boa aderência do modelo à situação real deste outro conjunto de dados.

A pesquisa permitiu indicar que a metodologia aplicada permite escalabilidade dos modelos, através da criação de modelos por região, ou corredor de tráfego, pois permite o processamento em paralelo dos modelos.

Entendemos que existe a possibilidade de avanços qualitativos com os dois modelos estatísticos. Para o modelo STARIMA a melhoria pode ocorrer, por exemplo, através da criação de modelos específicos para cada período do dia, de modo a incluir a sazonalidade do fluxo nos diversos horários do dia. Para os modelos com Dados em Painel, a melhoria pode ocorrer através do aumento da base de dados, seja com maior número de vias com características diferentes, seja com mais dias no banco de treino.

Por fim, concluímos que os objetivos iniciais foram alcançados através da elaboração de uma metodologia de estimação de tráfego, viável à aplicação em tempo próximo ao real. Os modelos apresentados têm como base séries temporais e o aspecto espaço temporal. Os estudos com as matrizes espaciais propiciaram a determinação de modelos mais adequados às estimações do volume de tráfego do que os modelos utilizados até o momento nas aplicações com STARIMA. Os testes de uso dos modelos de estimação mostraram a possibilidade de uso para previsões de até 1 hora, no caso de fluxos de tráfego típicos.

BIBLIOGRAFIA

- ABOUDOLAS, K.; PAPAGEORGIOU, M.; KOSMATOPOULOS, E. Store-and-forward based methods for the signal control problem in large-scale congested urban road networks. **Transportation Research Part C: Emerging Technologies**, v. 17, n. 2, p. 163–174, 1 abr. 2009.
- AGHAMOHAMMADI, R.; LAVAL, J. A. Dynamic Traffic Assignment using the Macroscopic Fundamental Diagram: A Review of Vehicular and Pedestrian Flow Models. 2018.
- AHMED, F.; HAWAS, Y. E. A Threshold-Based Real-Time Incident Detection System for Urban Traffic Networks. **Procedia - Social and Behavioral Sciences**, v. 48, p. 1713–1722, 2012.
- ALLSTRÖM, A. et al. Hybrid Approach for Short-Term Traffic State and Travel Time Prediction on Highways. **Transportation Research Record: Journal of the Transportation Research**, v. 2554, p. 60–68, 2016.
- ALMEIDA, E. **Econometria Espacial Aplicada**. 1. ed. Campinas: Editora Alinea, 2012.
- ANSELIN, L. **Spatial Models in Econometric Research**. Chicago: [s.n.]. Disponível em: <<https://www.researchgate.net/publication/350327936>>. Acesso em: 20 jun. 2021.
- BALTAGI, B. H. **Econometric Analysis of Panel Data (Springer Texts in Business and Economics) - Sixth Edition**. [s.l.: s.n.].
- BALTAGI, B. H.; BRESSON, G.; PIROTTE, A. Panel unit root tests and spatial dependence. **Journal of Applied Econometrics**, v. 22, n. 2, p. 339–360, 1 mar. 2007.
- BARROS, O. M. DE. **CIMU - Integrated center for urban mobility São Paulo, Brazil**. 21st World Congress on Intelligent Transport Systems, ITSWC 2014: Reinventing Transportation in Our Connected World. **Anais...2014**.
- BASSANI, M. et al. The effects of road geometrics and traffic regulations on driver-preferred speeds in northern Italy. An exploratory analysis. **Transportation Research Part F: Traffic Psychology and Behaviour**, v. 25, n. PART A, p. 10–26, 1 jul. 2014.
- BASSANI, M. et al. Random Effect Models to Predict Operating Speed Distribution on Rural Two-Lane Highways. **Journal of Transportation Engineering**, v. 142, n. 6, p. 04016019, 22 fev. 2016a.
- BASSANI, M. et al. Night-time and daytime operating speed distribution in urban arterials. **Transportation Research Part F: Traffic Psychology and Behaviour**, v. 42, p. 56–69, 1 out. 2016b.
- BEN-AKIVA, M. et al. Real Time Simulation of Traffic Demand-Supply Interactions within DynaMIT. Em: [s.l.] Springer, Boston, MA, 2002. p. 19–36.
- BEN-ELIA, E.; SHIFTAN, Y. Which road do I take? A learning-based model of route-choice behavior with real-time information. **Transportation Research Part A: Policy and Practice**, v. 44, n. 4, p. 249–264, 1 maio 2010.
- BIZ, G. **Simulações de pesos espaciais para o modelo STARMA e aplicações**. Piracicaba: Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo, 1 ago. 2014.
- BORDONS, C.; CAMACHO, E. F. **Model Predictive Control**. [s.l.: s.n.].

- BROOKS, M. et al. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. **The R Journal**, v. 9, n. 2, p. 378–400, 2017.
- BURGHOUT, W.; KOUTSOPOULOS, H. N.; ANDREASSON, I. Incident Management and Traffic Information Tools and Methods for Simulation-Based Traffic Prediction. **Transportation Research Record**, n. 2161, p. 20–28, 2010.
- CAI, P. et al. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. **Transportation Research Part C: Emerging Technologies**, v. 62, p. 21–34, 1 jan. 2016.
- CASTILLO, E. et al. A State-of-the-Art Review of the Sensor Location, Flow Observability, Estimation, and Prediction Problems in Traffic Networks. **Journal of Sensors**, v. 2015, p. 1–26, 29 out. 2015.
- CHEN, C. et al. The retrieval of intra-day trend and its influence on traffic prediction. **Transportation Research Part C: Emerging Technologies**, v. 22, p. 103–118, 1 jun. 2012.
- CHEN, M.; CHIEN, S. I. J. Dynamic Freeway Travel-Time Prediction with Probe Vehicle Data. **Transportation Research Record**, v. 1768, p. 157–161, 2001.
- CHENG, T. et al. **Modelling Dynamic Space-Time Autocorrelations of Urban Transport Network**. 11th international conference on geo- computation. **Anais...**2011.
- CHENG, T. et al. A Dynamic Spatial Weight Matrix and Localized Space-Time Autoregressive Integrated Moving Average for Network Modeling. **Geographical Analysis**, v. 46, n. 1, p. 75–97, 1 jan. 2014.
- CLIFF, A. D.; ORD, J. K. Space-Time Modelling with an Application to Regional Forecasting. **Transactions of the Institute of British Geographers**, v. 64, n. 64, p. 119, 1975.
- CONSELHO NACIONAL DE TRÂNSITO CONTRAN. **Manual Brasileiro de Sinalização de Trânsito VOLUME V Sinalização Semafórica**. Brasília, 2014.
- DARAGHMI, Y.-A.; YI, C.-W.; CHIANG, T.-C. Negative Binomial Additive Models for Short-Term Traffic Flow Forecasting in Urban Areas. **IEEE Transactions on Intelligent Transportation Systems**, v. 15, n. 2, p. 784–793, abr. 2014.
- DARDOR, M.; CHLYAH, M.; BOUMHIDI, J. **Incident Detection in signaled urban roads based on Genetic Algorithm and Support Vector Machine**. International Conference on Intelligent Systems and Computer Vision (ISCV). **Anais...**2018. Disponível em: <<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8354029>>. Acesso em: 5 nov. 2021
- DE LUCA, G.; GALLO, M. **Artificial neural networks for forecasting user flows in transportation networks: Literature review, limits, potentialities and open challenges**. 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017 - Proceedings. **Anais...**2017.
- DIA, H.; THOMAS, K. Development and evaluation of arterial incident detection models using fusion of simulated probe vehicle and loop detector data. **Information Fusion**, v. 12, n. 1, p. 20–27, 1 jan. 2011.
- DIAMANTOPOULOS, T. et al. **Investigating the effect of global metrics in travel time forecasting**. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. **Anais...**2013.
- DING, Q. Y. et al. Forecasting Traffic Volume with Space-Time ARIMA Model. **Advanced Materials Research**, v. 156–157, p. 979–983, 2011.

DNIT MINISTÉRIO DOS TRANSPORTES DEPARTAMENTO NACIONAL DE INFRA-ESTRUTURA DE TRANSPORTES DIRETORIA DE PLANEJAMENTO E PESQUISA COORDENAÇÃO-GERAL DE ESTUDOS E PESQUISA INSTITUTO DE PESQUISAS RODOVIÁRIAS. **MANUAL DE ESTUDOS DE TRÁFEGO**. [s.l: s.n.]. . Acesso em: 15 ago. 2020.

DUAN, P. et al. **STARIMA-based Traffic Prediction with Time-varying Lags**. IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). **Anais...IEEE**, nov. 2016.

DUAN, P. et al. A Unified Spatio-Temporal Model for Short-Term Traffic Flow Prediction. **IEEE Transactions on Intelligent Transportation Systems**, v. 20, n. 9, p. 3212–3223, 1 set. 2019.

ELHORST, J. P. Specification and Estimation of Spatial Panel Data Models. **International Regional Science Review**, v. 26, n. 3, p. 244–268, 26 jul. 2003.

ELHORST, J. P. **Spatial Panel Data Analysis**. [s.l.] Springer International Publishing, 2017. Disponível em: <www.regroningen.nl/elhorst>. Acesso em: 24 out. 2020.

ELHORST, J. P.; HALLECK VEGA, S. **The SLX model: Extensions and the sensitivity of spatial spillovers to W**. [s.l: s.n.]. Disponível em: <<http://www.rug.nl/research/portal>>. Acesso em: 24 out. 2020.

ERMAGUN, A.; LEVINSON, D. Spatiotemporal traffic forecasting: review and proposed directions. **Transport Reviews**, v. 38, n. 6, p. 786–814, 2 nov. 2018.

ERMAGUN, A.; LEVINSON, D. Spatiotemporal short-term traffic forecasting using the network weight matrix and systematic detrending. **Transportation Research Part C: Emerging Technologies**, v. 104, p. 38–52, 1 jul. 2019.

FÁVERO, L. P. L.; BELFIORE, P. **Manual de Análise de Dados - Estatística e Modelagem Multivariada com Excel, SPSS e Stata**. 3. ed. Rio de Janeiro: Elsevier Ltd, 2017.

FENG, L. et al. Geographically and temporally weighted neural network for winter wheat yield prediction. **Remote Sensing of Environment**, v. 262, p. 112514, 1 set. 2021.

FHWA OFFICE OF HIGHWAY INFORMATION MANAGEMENT. **ITS PROBE VEHICLE TECHNIQUES**. [s.l: s.n.].

GANG, C.; SHOUHUI, W.; XIAOBO, X. **Review of spatio-temporal models for short-term traffic forecasting**. 2016 IEEE International Conference on Intelligent Transportation Engineering (ICITE). **Anais...IEEE**, ago. 2016.

GENTILE, G.; MESCHINI, L. **Using dynamic assignment models for real-time traffic forecast on large urban networks**. In Proceedings of the 2nd International Conference on Models and Technologies for Intelligent Transportation Systems. **Anais...2011**.

GENTILI, M.; MIRCHANDANI, P. B. Review of optimal sensor location models for travel time estimation. **Transportation Research Part C: Emerging Technologies**, v. 90, p. 74–96, 1 maio 2018.

GEORGESCU, L.; ZEITLER, D.; STANDRIDGE, C. R. Intelligent transportation system real time traffic speed prediction with minimal data. **Journal of Industrial Engineering and Management**, v. 5, n. 2, p. 431–441, 29 dez. 2012.

GIACOMO COMO. On resilient control of dynamical flow networks. **Annual Reviews in Control**, v. 43, p. 80–90, 1 jan. 2017.

GLIŠOVIĆ, N. et al. A hybrid model for forecasting the volume of passenger flows on Serbian railways. **Operational Research**, v. 16, n. 2, 2016.

GONÇALVES, E. S. **Análise de padrões operacionais da frota de ônibus de transporte público no município de São Paulo e a influência de fatores climáticos em sua dinâmica**. São Paulo: Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo, 13 mar. 2018.

GRIFFITH, D. A. Modeling spatio-temporal relationships: Retrospect and prospect. **Journal of Geographical Systems**, v. 12, n. 2, p. 111–123, 30 abr. 2010.

GRUBBS, F. E.; GRUBBS, F. E.; GRUBBS, F. E. Procedures for Detecting Outlying Observations in Samples. **Technometrics**, v. 11, n. 1, p. 1–21, 1 fev. 1969.

GUAN, D.; HUANG, L.; QU, Q. **A Predicting Method of Urban Traffic Network Volume Based on STARIMA Model**. CICTP 2017. **Anais...**Reston, VA: American Society of Civil Engineers, 18 jan. 2018.

GUO, J.; HUANG, W.; WILLIAMS, B. M. Real time traffic flow outlier detection using short-term traffic conditional variance prediction. **Transportation Research Part C: Emerging Technologies**, v. 50, p. 160–172, 1 jan. 2015.

HABTEMICHAEL, F. G.; CETIN, M. Short-term traffic flow rate forecasting based on identifying similar traffic patterns. **Transportation Research Part C: Emerging Technologies**, v. 66, 2016.

HIRIOTAPPA, K. et al. A streaming algorithm for online estimation of temporal and spatial extent of delays. **Journal of Advanced Transportation**, v. 2017, 2017.

HOU, T. et al. Calibration of Traffic Flow Models under Adverse Weather and Application in Mesoscopic Network Simulation Procedures. **Transportation Research Record**, n. November 2012, p. 23, 2013.

ISLAM, M. T.; EL-BASYOUNY, K. Multilevel models to analyze before and after speed data. **Analytic Methods in Accident Research**, v. 8, p. 33–44, 1 dez. 2015.

ITABORAI, P. V. **Itaborala/SP_Traffic_Dataset_Builder**. , 2022. Disponível em: https://github.com/Itaborala/SP_Traffic_Dataset_Builder, <https://doi.org/10.5281/zenodo.7255096>.

KAIN, M. P.; BOLKER, B. M.; MCCOY, M. W. A practical guide and power analysis for GLMMs: Detecting among treatment variation in random effects. **PeerJ**, v. 2015, n. 9, p. e1226, 17 set. 2015.

KAMARIANAKIS, Y. **SPATIAL-TIME SERIES MODELING: A REVIEW OF THE PROPOSED METHODOLOGIES**. [s.l: s.n.].

KAMARIANAKIS, Y.; PRASTACOS, P. **Forecasting Traffic Flow Conditions in an Urban Network: Comparison of Multivariate and Univariate Approaches**. **Transportation Research Record. Anais...**National Research Council, 2003.

KAMARIANAKIS, Y.; PRASTACOS, P. Space-time modeling of traffic flow. **Computers and Geosciences**, v. 31, n. 2, p. 119–133, 1 mar. 2005.

KANNAN, S. K. Comparison of methods for detecting outliers. **International Journal of Scientific & Engineering Research**, v. 4, n. 9, 2013.

KARLAFTIS, M. G.; VLAHOIANNI, E. I. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. **Transportation Research Part C: Emerging Technologies**, v. 19, n. 3, p. 387–399, 2011.

KIM, J. Does roadwork improve road speed? Evidence from urban freeways in California. **Regional Science and Urban Economics**, v. 93, p. 103773, 1 mar. 2022.

KIM, K. et al. Dynamic Vehicular Route Guidance Using Traffic Prediction Information. **Mobile Information Systems**, v. 2016, p. 1–11, 4 jul. 2016.

KUCHARSKI, R.; KOSTIC, B.; GENTILE, G. **Real-time traffic forecasting with recent DTA methods**. 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). **Anais...IEEE**, jun. 2017.

LAÑA, I. et al. Road Traffic Forecasting : Recent Advances and New Challenges. **IEEE Intelligent Transportation Systems Magazine**, n. April, p. 93–109, 2018.

LI, L. et al. **Traffic prediction, data compression, abnormal data detection and missing data imputation: An integrated study based on the decomposition of traffic time series**. 2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014. **Anais...2014**.

LI, L. et al. Robust causal dependence mining in big data network and its application to traffic flow predictions. **Transportation Research Part C: Emerging Technologies**, v. 58, p. 292–307, 1 set. 2015.

LI, L.; LI, Y.; LI, Z. Efficient missing data imputing for traffic flow by considering temporal and spatial dependence. **Transportation Research Part C: Emerging Technologies**, v. 34, p. 108–120, 1 set. 2013.

LIN, S. et al. Efficient network-wide model-based predictive control for urban traffic networks. **Transportation Research Part C: Emerging Technologies**, v. 24, p. 122–140, 1 out. 2012.

LIN, S. L. et al. The Application Of Space-Time ARIMA Model On Traffic Flow Forecasting. **Proceedings of the 2009 International Conference on Machine Learning and Cybernetics**, v. 6, p. 3408–3412, 2009.

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, jan. 2005.

LUKUSA, M. T.; HING PHOA, F. K. A Horvitz-type estimation on incomplete traffic accident data analyzed via a zero-inflated Poisson model. **Accident Analysis and Prevention**, v. 134, 1 jan. 2020.

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 & Prevention**, v. 40, n. 3, p. 964–975, 1 maio 2008.

MAERIVOET, S.; DE MOOR, B. Traffic Flow Theory. 15 jul. 2005.

METRÔ SP. **Metrô São Paulo | Pesquisa Origem Destino - 50 anos**. Disponível em: <<http://www.metro.sp.gov.br/pesquisa-od/pesquisa-od.aspx>>. Acesso em: 3 fev. 2018.

MIN, W.; WYNTER, L. Real-time road traffic prediction with spatio-temporal correlations. **Transportation Research Part C: Emerging Technologies**, v. 19, n. 4, p. 606–616, 1 ago. 2011.

MIN, X. et al. **Short-term traffic flow forecasting of urban network based on dynamic STARIMA model**. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. **Anais...2009**.

MIN, X.; HU, J.; ZHANG, Z. Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model. **IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC**, p. 1535–1540, 2010.

MITSAKIS, E.; SALANOVA, J. M.; GIANNOPOULOS, G. Combined dynamic traffic assignment and urban traffic control models. **Procedia - Social and Behavioral Sciences**, v. 20, p. 427–436, 1 jan. 2011.

MUIIOZ, L. et al. **Traffic Density Estimation with the Cell Transmission Modell**. Proceedings of the American Control Conference. **Anais...**Denver: 2003. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1240418>>. Acesso em: 28 out. 2017

OH, S. et al. Short-term travel-time prediction on highway: A review on model-based approach. **KSCE Journal of Civil Engineering**, p. 1–13, 5 maio 2017.

OTOSHI, T. et al. Hierarchical Model Predictive Traffic Engineering. **IEEE/ACM Transactions on Networking**, v. 26, n. 4, p. 1754–1767, 1 ago. 2018.

PAN, B. et al. **Forecasting Spatiotemporal Impact of Traffic Incidents on Road Networks**. 2013 IEEE 13th International Conference on Data Mining. **Anais...**IEEE, dez. 2013. Disponível em: <<http://ieeexplore.ieee.org/document/6729543/>>. Acesso em: 18 nov. 2018

PARKANY, E.; XIE, C. **A Complete Review of Incident Detection Algorithms & Their Deployment: What Works and What Doesn't**. [s.l: s.n.]. Disponível em: <<http://docs.trb.org/00988875.pdf>>. Acesso em: 8 jul. 2020.

PAVLYUK, D. **Short-term Traffic Forecasting Using Multivariate Autoregressive Models**. Procedia Engineering. **Anais...**Elsevier Ltd, 1 jan. 2017. Disponível em: <www.sciencedirect.com>. Acesso em: 5 jul. 2019

PAVLYUK, D. Feature selection and extraction in spatiotemporal traffic forecasting: a systematic literature review. **European Transport Research Review**, v. 11, n. 1, p. 6, 24 dez. 2019.

PAVLYUK, D. Temporal aggregation effects in spatiotemporal traffic modelling. **Sensors (Switzerland)**, v. 20, n. 23, p. 1–21, 1 dez. 2020.

PAZ, K. D.; TRINTINALIA, M. M. T. **STARIMA**. Disponível em: <<https://starima.shinyapps.io/starima/>>. Acesso em: 22 jun. 2021.

PELL, A.; MEINGAST, A.; SCHAUER, O. Trends in Real-time Traffic Simulation. **Transportation Research Procedia**, v. 25, p. 1477–1484, 1 jan. 2017.

PFEIFER, P. E.; BODILY, S. E. A Test of Space-Time ARMA Modelling and Forecasting of Hotel Data. **Journal of Forecasting**, v. 9, p. 255–272, 1990.

PFEIFER, P. E.; DEUTSCH, S. J. A STARIMA Model-Building Procedure with Application to Description and Regional Forecasting. **Transactions of the Institute of British Geographers**, v. 5, n. 3, p. 330, 1980a.

PFEIFER, P. E.; DEUTSCH, S. J. **A Three-Stage Iterative Procedure for Space-Time Modeling****TECHNOMETRICS**. [s.l: s.n.]. . Acesso em: 3 abr. 2021b.

PFEIFER, P. E.; DEUTSCH, S. J. Seasonal Space-Time ARIMA Modeling. **Geographical Analysis**, v. 13, n. 2, p. 117–133, 1981.

PORTUGAL, L. DA S. **Simulação de Tráfego: Conceitos e Técnicas de Modelagem**. [s.l.] Editora Interciência Ltda., 2005.

PUANGNAK, K.; CHIVAPREECHA, S. **A review study of incident detection algorithms with performance index parameter**. Proceedings of the 16th International Conference on Electrical

- Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2019. **Anais...**Institute of Electrical and Electronics Engineers Inc., 1 jul. 2019. . Acesso em: 8 jul. 2020
- QIU, T. Z. et al. Estimation of Freeway Traffic Density with Loop Detector and Probe Vehicle Data. **Transportation Research Record: Journal of the Transportation Research Board**, v. 2178, p. 21–29, 2010.
- RAO, T. S.; ANTUNES, A. M. C. SPATIO-TEMPORAL MODELLING OF TEMPERATURE DATA. **The IMA Volumes in Mathematics and its Applications**, v. 139, p. 260, 2004.
- RONCOLI, C.; PAPAMICHAIL, I.; PAPAGEORGIOU, M. Hierarchical model predictive control for multi-lane motorways in presence of Vehicle Automation and Communication Systems. **Transportation Research Part C: Emerging Technologies**, v. 62, p. 117–132, 1 jan. 2016.
- ROSNER, B. Percentage Points for a Generalized ESD Many-Outlier Procedure. **Technometrics**, v. 25, n. 2, p. 165, maio 1983.
- SAHA, P. et al. Application of Principal Component Analysis for Outlier Detection in Heterogeneous Traffic Data. **Procedia Computer Science**, v. 83, p. 107–114, 1 jan. 2016.
- SALAMANIS, A. et al. Evaluating the Effect of Time Series Segmentation on STARIMA-Based Traffic Prediction Model. **IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC**, v. 2015- Octob, p. 2225–2230, 30 out. 2015.
- SALAMANIS, A. et al. Managing Spatial Graph Dependencies in Large Volumes of Traffic Data for Travel-Time Prediction. **IEEE Transactions on Intelligent Transportation Systems**, v. 17, n. 6, p. 1678–1687, 1 jun. 2016.
- SCHIMBINSCHI, F. et al. Topology-regularized universal vector autoregression for traffic forecasting in large urban areas. **Expert Systems with Applications**, v. 82, p. 301–316, 1 out. 2017.
- SECRETARIA MUNICIPAL DE TRANSPORTES. **Termo de Referência do Edital 002/SMT/2019**. , 2019.
- SEO, T. et al. Traffic state estimation on highway: A comprehensive survey. **Annual Reviews in Control**, v. 43, p. 128–151, 1 jan. 2017.
- SERNEELS, S.; VERDONCK, T. Principal component analysis for data containing outliers and missing elements. **Computational Statistics & Data Analysis**, v. 52, n. 3, p. 1712–1727, 1 jan. 2008.
- SHI, Y. et al. Detecting anomalies in spatio-temporal flow data by constructing dynamic neighbourhoods. . 1 jan. 2018, p. 80–96.
- SLOBODEN, J. et al. **Guidebook on the Utilization of Dynamic Traffic Assignment in Modeling**. [s.l: s.n.]. Disponível em: <<https://ops.fhwa.dot.gov/publications/fhwahop13015/fhwahop13015.pdf>>. Acesso em: 13 jan. 2018.
- STATHOPOULOS, A.; KARLAFTIS, M. **Temporal and Spatial Variations of Real-Time Traffic Data in Urban Areas**. [s.l: s.n.].
- STATHOPOULOS, A.; KARLAFTIS, M. G. A multivariate state space approach for urban traffic flow modeling and prediction. **Transportation Research Part C: Emerging Technologies**, v. 11, n. 2, p. 121–135, 1 abr. 2003.
- T. DARWISH; K. ABU BAKAR. Traffic density estimation in vehicular ad hoc networks: A review. **Ad Hoc Networks**, v. 24, p. 337–351, 1 jan. 2015.

- TANG, L.; THAKURIAH, P. V. Ridership effects of real-time bus information system: A case study in the City of Chicago. **Transportation Research Part C: Emerging Technologies**, v. 22, p. 146–161, 1 jun. 2012.
- TAO, Y.; SUN, P.; BOUKERCHE, A. A Novel Travel-Delay Aware Short-Term Vehicular Traffic Flow Prediction Scheme for VANET. **IEEE Wireless Communications and Networking Conference, WCNC**, v. 2019- April, 1 abr. 2019.
- TASCIKARAOGLU, A. Evaluation of spatio-temporal forecasting methods in various smart city applications. **Renewable and Sustainable Energy Reviews**, v. 82, 2018.
- TOBLER, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. **Economic Geography, Supplement : Proceedings International Geographical Union Commission on Quantitative Methods**, v. 46, p. 234–240, 1970.
- TRANSPORTATION RESEARCH BOARD. **Highway Capacity Manual, Sixth Edition: A Guide for Multimodal Mobility Analysis**. 6. ed. Washington: [s.n.].
- TREMBLAY, J. M.; CIRILLO, C.; BASSANI, M. Updating and transferring Random Effect models: The case of operating speed percentile estimation. **Transportation Research Part A: Policy and Practice**, v. 148, p. 286–304, 1 jun. 2021.
- TSELENTIS, D. I.; VLAHOGIANNI, E. I.; KARLAFTIS, M. G. Improving short-term traffic forecasts: To combine models or not to combine? **IET Intelligent Transport Systems**, v. 9, n. 2, 2015.
- VAN LINT, J. W. C.; HOOGENDOORN, S. P.; VAN ZUYLEN, H. J. Accurate freeway travel time prediction with state-space neural networks under missing data. **Transportation Research Part C: Emerging Technologies**, v. 13, n. 5–6, p. 347–369, 1 out. 2005.
- VIEGO, V. N.; VOLONTÉ, C. B. Estimaciones del tráfico urbano de pasajeros de colectivos de Bahía Blanca, Argentina. 2007-2014. **Economía Sociedad y Territorio**, v. xvii, n. 1, p. 561–580, 2 set. 2017.
- VLAHOGIANNI, E. I. **Computational intelligence and optimization for transportation big data: Challenges and opportunities**. [s.l: s.n.]. v. 38
- VLAHOGIANNI, E. I.; KARLAFTIS, M. G.; GOLIAS, J. C. Short-term traffic forecasting: Where we are and where we're going. **Transportation Research Part C: Emerging Technologies**, v. 43, n. Part C, p. 3–19, 1 jun. 2014.
- WADDOO, S. A. et al. **Prediction of traffic density from wireless cellular data**. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). **Anais...IEEE**, nov. 2016. Disponível em: <<http://ieeexplore.ieee.org/document/7795611/>>. Acesso em: 3 fev. 2018
- WANG, J.; TSAPAKIS, I.; ZHONG, C. A space-time delay neural network model for travel time prediction. **Engineering Applications of Artificial Intelligence**, v. 52, p. 145–160, 1 jun. 2016.
- WANG, S.; MAO, G. Missing Data Estimation for Traffic Volume by Searching an Optimum Closed Cut in Urban Networks. **IEEE Transactions on Intelligent Transportation Systems**, v. 20, n. 1, p. 75–86, jan. 2019.
- WANG, X.; GUO, Q.; TARKO, A. P. Modeling speed profiles on mountainous freeways using high resolution data. **Transportation Research Part C: Emerging Technologies**, v. 117, p. 102679, 1 ago. 2020.

- WANG, Y. et al. Dynamic traffic assignment: A review of the methodological advances for environmentally sustainable road transportation applications. **Transportation Research Part B**, v. 111, p. 370–394, 2018.
- WANG, Y.; PAPAGEORGIOU, M.; MESSMER, A. RENAISSANCE - A unified macroscopic model-based approach to real-time freeway network traffic surveillance. **Transportation Research Part C: Emerging Technologies**, v. 14, n. 3, p. 190–212, 1 jun. 2006.
- WANG, Z.; QI, X.; JIANG, H. Estimating the spatiotemporal impact of traffic incidents: An integer programming approach consistent with the propagation of shockwaves. **Transportation Research Part B: Methodological**, v. 111, p. 356–369, 1 maio 2018.
- WILRICH, P.-T. Critical values of Mandel's h and k, the Grubbs and the Cochran test statistic. **AStA Advances in Statistical Analysis** 2011 97:1, v. 97, n. 1, p. 1–10, 7 dez. 2011.
- WU, Y.-J. J. et al. Urban Traffic Flow Prediction Using a Spatio-Temporal Random Effects Model. **Journal of Intelligent Transportation Systems**, v. 20, n. 3, p. 282–293, 2015.
- XIONG, Z. et al. Intelligent transportation systems for smart cities: a progress review. **Science China Information Sciences**, v. 55, n. 12, p. 2908–2914, 24 dez. 2012.
- XU, Y.; KONG, Q.-J.; LIU, Y. **Comparison of urban traffic prediction methods between UTN-based spatial model and time series models**. 2012 15th International IEEE Conference on Intelligent Transportation Systems. **Anais...Anchorage: IEEE**, set. 2012.
- YANG, J. et al. Grey coupled prediction model for traffic flow with panel data characteristics. **Entropy**, v. 18, n. 12, 2016.
- YE, B. L. et al. A hierarchical model predictive control approach for signal splits optimization in large-scale urban road networks. **IEEE Transactions on Intelligent Transportation Systems**, v. 17, n. 8, p. 2182–2192, 1 ago. 2016.
- ZHAN, X. et al. Multi-step-ahead traffic speed forecasting using multi-output gradient boosting regression tree. **Journal of Intelligent Transportation Systems: Technology, Planning, and Operations**, v. 24, n. 2, p. 125–141, 3 mar. 2019.
- ZHAO, Y. et al. A new Seasonal Difference Space-Time Autoregressive Integrated Moving Average (SD-STARIMA) model and spatiotemporal trend prediction analysis for Hemorrhagic Fever with Renal Syndrome (HFRS). **PLoS ONE**, v. 13, n. 11, 1 nov. 2018.
- ZHOU, C.; WANG, S. Examining the determinants and the spatial nexus of city-level CO₂ emissions in China: A dynamic spatial panel analysis of China's cities. **Journal of Cleaner Production**, v. 171, p. 917–926, 10 jan. 2018.
- ZHOU, Z. et al. Two-Level Hierarchical Model-Based Predictive Control for Large-Scale Urban Traffic Networks. **IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY**, v. 25, n. 2, 2017.

APÊNDICES

A. Trabalhos relacionados e decorrentes:

a. Artigo em fase de revisão em Revista Científica

O artigo MATRIZ ESPAÇO-TEMPORAL NA PREVISÃO DE TRÁFEGO DE CURTO PRAZO BASEADA EM SÉRIES TEMPORAIS DE DADOS foi escrito em coautoria com os professores está em fase de revisão para envio a revista científica, visando sua publicação. Parte do estudo apresentado no artigo está incluída nesta tese.

a. Artigo científico em Congresso

O artigo APLICAÇÃO DO MODELO STARIMA EM PREVISÃO DE TRÁFEGO DE CURTO PRAZO: UM ESTUDO DE CASO NA CIDADE DE SÃO PAULO foi escrito com a coautoria das autoras do aplicativo STARIMA em R-Shiny, Kaiene Domingues Paz e Michelle Mieko Taniguchi Trintinalia, e do Orientador Prof. Dr. Claudio Luiz Marte. Artigo completo aceito para apresentação no 36º Congresso da ANPET – Fortaleza-CE, 11/2022.

b. Trabalhos de Conclusão de Curso (TCC)

Os trabalhos de conclusão de curso TCCs descritos a seguir utilizaram os dados das séries temporais pontuais preparadas na primeira fase desta pesquisa e foram coorientados pelo doutorando através de reuniões on-line, fornecimento de bibliografia e apoio na análise dos resultados parciais e finais.

i. ARIMA

Título: Aplicação de ARIMA em Previsão de Tráfego de Curto Prazo

Autores: Domenico Zemma, Marcella Gomes da Costa Pereira e Yuri Gil Ferreira

Foi avaliado o uso do modelo univariado ARIMA para a previsão de tráfego de curto prazo, tendo sido desenvolvidos scripts em Python de aplicação do modelo.

ii. STARIMA

Título: Aplicação de STARIMA em Previsão de Tráfego de Curto Prazo Um Estudo de Caso na Cidade de São Paulo

Autores: Kaiene Domingues Paz, Marco Aurélio Tavares Marques Da Silva e Michelle Mieko Taniguchi Trintinalia

Foi avaliado o uso do modelo STARIMA na previsão de tráfego de curto prazo, com desenvolvimento de aplicativo em R no pacote de expansão R Shiny, conforme descrito no Apêndice D.

iii. STARIMA com Diferenciação Sazonal

Título: Aplicação de STARIMA Sazonal para Previsão de Tráfego de Curto Prazo – Análise de Caso no Corredor Norte Sul da Cidade de São Paulo

Autores: Mariane Ayumi Bando e Ricardo Bonfiglioli Junqueira Andrade

Foi avaliada a aplicação de diferenciação sazonal para os dados tratados no modelo STARIMA. Os resultados comparativos indicam que o uso da sazonalidade diferenciada em uma semana não gerou estimações mais acuradas em relação ao mesmo modelo sem essa diferenciação.

iv. Random Forest

Título: Aplicação do método estatístico de árvores de regressão em transportes

Autores: Álvaro Miguel Linares, Caio Gonçalves de Moraes e Caio Ubaldo Neri

Foi avaliado o uso de Random Forest na previsão de tráfego de um único radar. Neste caso foram utilizadas ferramentas já disponíveis na plataforma Python.

v. GBRT

Título: Previsão de tráfego a curto prazo utilizando LightGBM com dados de radares

Autor: Murilo Yoshio Matsubara

O estudo comparou a aplicação de LightGBM e de Random Forest na previsão de curto prazo em um sensor, utilizando dados dos sensores anterior e posterior e dados defasados

em 15 intervalos à frente e atrás. O estudo concluiu que o LithtGBM teve desempenho superior.

c. Iniciações Científicas (IC)

vi. Guilherme Mazzuia Stocco

Desenvolvimento de scripts para o agrupamento dos dados dos sensores, sob coordenação do doutorando, e elaboração de laboratórios para uso do modelo ARIMA em aulas da graduação e pós-graduação

vii. Rodrigo Pina Bluma

Desenvolvimento de script para a determinação do tempo e velocidade de percurso nos trechos, sob coordenação do doutorando, e elaboração de laboratórios para uso do modelo STARIMA em aulas da graduação e pós-graduação

viii. Emerson Fernandes Sabino

Atividades em Python de desenvolvimento das novas bases de dados, e elaboração de laboratório para uso do modelo STARIMA com Diferenciação Sazonal

B. Aplicação de Dados em Painel na predição de volumes

Neste Apêndice, apresentamos resumidamente a aplicação de Regressão com Painel de Dados com característica aninhada e não linear, correspondendo ao HNM apresentado no item 3.2.2.

Esta aplicação foi colocada neste apêndice, pois foi desenvolvida em paralelo às pesquisas da tese e se baseou em dados diferentes dos que foram tratados ao longo da tese. A aplicação foi feita a fim de aplicar a metodologia, em caso real, onde a equipe de planejamento da CET necessitava complementar dados volumétricos de alguns pontos da cidade.

Durante os anos 2020 e 2021 a pandemia impossibilitou que os levantamentos volumétricos anuais fossem efetuados da maneira usual, qual seja, através de contagem manual com equipe em campo. Para suprir tal deficiência, está em andamento estudo de determinação desses valores com base nos registros dos equipamentos de fiscalização (radares) existentes em locais próximos de onde eram efetuadas as contagens nos anos anteriores. Ocorre que alguns equipamentos deixaram de efetuar as contagens por períodos longos, acima de um dia, e os dados precisavam ser complementados.

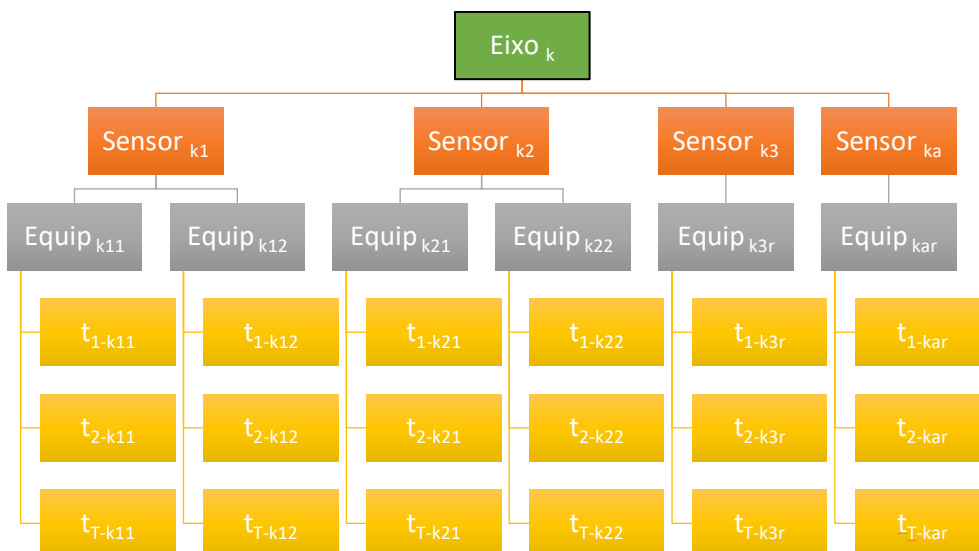
Assim, esta aplicação teve como objetivo complementar os dados dos radares com base nos dados históricos disponíveis de cada local e nos equipamentos vizinhos. O ano de 2019, foi escolhido para o estudo inicial, pois seria possível comparar os resultados com os valores obtidos nas contagens manuais efetuadas naquele ano.

Apenas os dados de volume estavam disponíveis para alguns locais selecionados pela equipe antes do estudo, de modo que embora existam diversos equipamentos numa mesma via, somente os equipamentos próximos aos locais históricos de contagem foram selecionados. No total foram selecionados 65 equipamentos, que correspondiam a 33 locais de fiscalização (sensores), distribuídos por 5 corredores de tráfego: Av. Rebouças/R. da Consolação, Av. Brasil, Av. 9 de Julho, Corredor Norte Sul e Av. Faria Lima. A Figura B-1 apresenta a hierarquização utilizada na construção do modelo.

Os dados disponíveis incluíam os 7 primeiros dias de todos os 12 meses de 2019. No total estavam disponíveis 131040 registros. Os primeiros 9 meses foram utilizados na base de treino e os 3 últimos na base de testes.

Foi desenvolvido um modelo HNM de 4 níveis, no qual, para os efeitos fixos foram utilizadas as variáveis de volume, volume defasado em até 8 intervalos, período do dia (5 períodos), dia da semana (7 dias) e dia útil (2 classes). Foram testadas variáveis que unificavam a característica período e dia da semana (35 classes). Na próxima etapa do estudo serão aplicadas variáveis relativas à distância entre os sensores e s variáveis de efeito aleatório.

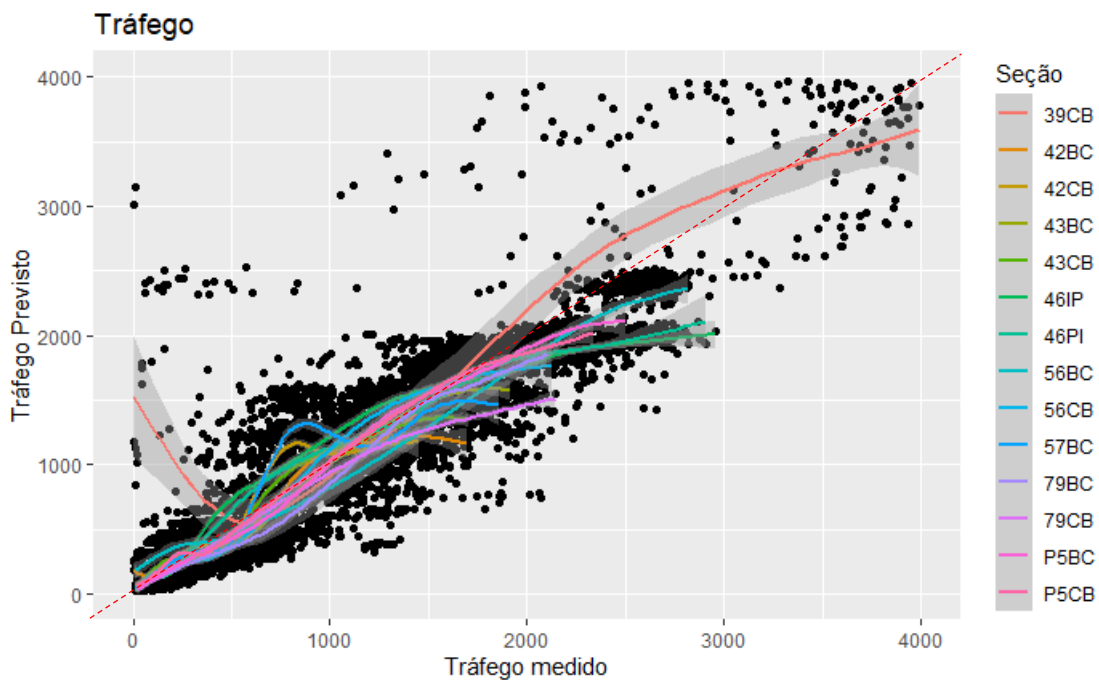
Figura B-1 – Hierarquização aplicada no modelo



Fonte: o autor

Os resultados obtidos podem ser ilustrados pela Figura B-2, que apresenta a relação entre valores medidos e previstos, com indicação das seções avaliadas.

Figura B-2 – Previsto x Medido – Modelo HNM4



C. Scripts

Neste apêndice são apresentados os algoritmos especialmente desenvolvidos para esta pesquisa, salientando que este trabalho contou com a colaboração de alunos da graduação do curso de Engenharia Civil da Escola Politécnica da USP e de um ex-aluno da Faculdade de Física da USP. Os scripts dos módulos 2 e 4 estão disponíveis para acesso no google/drive:

<https://drive.google.com/drive/folders/1PKlxNYgC91LllkfXjkSfAhonAFDHWLcu?usp=sharing> ;

a) Scripts do Módulo 1 e Tratamento de Dados Sensíveis:

Os scripts do Módulo 1 e de tratamento dos Dados Sensíveis foram elaborados por Paulo Vitor Itaborai de Barros (ex aluno da Física/USP), como parte de seus estudos sobre a linguagem Python e de processamento de dados em nuvem (ITABORAI, 2022).

b) Scripts do Módulo 2:

A construção inicial dos scripts para a montagem das séries temporais do Módulo 2 foi feita por Guilherme Mazzuia Stocco (Dados Pontuais), Rodrigo Pina Bluma (Dados de Trecho) e Emerson Fernandes Sabino (unificação dos dados Pontuais e de Trecho) durante atividades de Iniciação Científica (IC) na Escola Politécnica da USP, com base em estrutura proposta por este autor. Posteriormente, foram necessários ajustes decorrentes de novas necessidades surgidas durante os estudos, o que foi feito pelo próprio autor.

c) Scripts do Módulo 4:

Para o Módulo 4 temos que o script de Estimação com HLM foi elaborado pelo autor com base na biblioteca glmmTMB do R (BROOKS et al., 2017).

Na estimação baseada em STARIMA foi utilizada a ferramenta STARIMA descrita no Apêndice D.

D. Aplicativo STARIMA no R Shiny

O aplicativo utilizado para a execução dos experimentos foi desenvolvido durante o TCC das autoras e de Marco Aurélio Tavares Marques da Silva, utilizando a linguagem R, no pacote de expansão Shiny e permite efetuar previsões de curto prazo, sendo possível definir o horizonte de previsão e a execução de previsões com ou sem a rolagem dos dados. O acesso ao aplicativo é feito pela internet no site <https://starima.shinyapps.io/starima/> (PAZ; TRINTINALIA, 2021).

Em sua implementação, utiliza o conceito do Teste Aumentado de Dick Fuller para a determinação da ordem d de diferenciação do modelo (KAMARIANAKIS; PRASTACOS, 2005); o método de Levenberg-Marquadt para a resolução de sistemas não lineares, que permite a definição dos parâmetros alfas e betas – utilizando a suposição de ruído branco dos dados temporais, dispensando assim a necessidade do *back-forecasting* – pois no próprio ambiente computacional RStudio já existe uma biblioteca com essa função.

Os parâmetros Coeficiente de Eficácia, MAPE, RMSE, MAE, RSS e Box-Pierce são disponibilizados para a calibração do modelo e análise de acurácia. O valor do índice de Box-Pierce está parametrizado no aplicativo para ser calculado com base nos 25 intervalos anteriores ao momento do corte de treino, de modo que o valor máximo de graus de liberdade é definido descontando-se a quantidade de parâmetros do modelo.

Vale ressaltar ainda que cada tipo de previsão acompanha seu respectivo intervalo de segurança, o que facilita a compreensão de até onde a eficácia do modelo é confiável. Essa incerteza da previsão é computada supondo que os resíduos gerados na previsão seguem distribuição gaussiana. O aplicativo está construído com um limite de 50 iterações para chegar no resultado final, sendo que após esse limite o resultado é apresentado mesmo sem que tenha ocorrido a convergência total. Modelos que não convergiram dentro das primeiras 50 iterações foram descartados.

O script deste aplicativo pode ser acessado em contato com as autoras em email disponível no site do aplicativo.

E. Detecção de incidentes

Muitos incidentes que ocorrem nas vias, inclusive acidentes de trânsito, deixam de ser informados aos órgãos de controle e não são detectados por equipes de campo ou sistemas de monitoramento e diversos esforços são feitos de modo a complementar dados e identificar a ocorrência desses fatos. Nessa direção, análise dos registros de acidentes em Changhua (Taiwan) com aplicação de regressão por Poisson Inflada de Zeros (ZIP) estimaram a ausência de informações de 30% dos incidentes (LUKUSA; HING PHOA, 2020). Em estudo relativo à Los Angeles (EUA) verificou-se a ausência de 55% dos dados na base oficial (PAN et al., 2013).

Vários sistemas de detecção de incidentes foram implementados em Centrais Operacionais dos Estados Unidos, a maior parte deles tratava de vias expressas e não apresentava boa acurácia, apesar disto esses sistemas continuaram sendo de grande importância aos gestores dessas centrais operacionais (PARKANY; XIE, 2005). Apesar da existência desse tipo de sistema, não foi possível confirmar sua implementação em alguma central operacional brasileira até a presente data.

Em análise comparativa das características do fluxo de tráfego (velocidade, volume, densidade), (PUANGNAK; CHIVAPREECHA, 2019) compararam 8 algoritmos de detecção de incidentes e concluíram que os algoritmos baseados em SARIMA, Algoritmo Bayesiano e inteligência artificial obtiveram índices de identificação de 99 a 100% e de falso alarme de 0 a 1,5%.

(DIA; THOMAS, 2011) utilizaram um modelo baseado em redes neurais, veículo-sonda e laços detectores nas entradas e saídas do trecho de estudo e obtiveram uma detecção de até 90% dos incidentes e apenas 0,5% de falso alarme. (DARDOR; CHLYAH; BOUMHIDI, 2018) utilizaram algoritmos genéticos e SVM para detectar incidentes em trecho único de via com 1 cruzamento e obtiveram detecção acima de 90% para a via arterial e 2,3% de falsos alarmes. Embora os resultados destes dois estudos sejam relevantes, deve-se considerar que a instalação de sensores em todos os acessos é impraticável em áreas urbanas complexas como a de São Paulo e outras grandes cidades, seja pelo custo de implantação e manutenção seja por sua operação.

(AHMED; HAWAS, 2012) analisaram o desvio (valor medido subtraído da média) da velocidade média e volume no local e a montante e jusante, considerando também a

velocidade média obtida apenas durante a fase verde do semáforo. Os modelos baseados em regressão linear múltipla para 4 faixas de volume de tráfego foram capazes de identificar 51% dos incidentes e geraram até 17% de falsos alarmes. A obtenção da velocidade média durante o verde do semáforo é possível de se obter em simuladores como o utilizado nesse estudo, todavia sua aplicação na prática implica na integração do sistema de semáforos com o sistema de sensores de velocidade.

Numa abordagem interessante, (GUO; HUANG; WILLIAMS, 2015) efetuaram a previsão utilizando um modelo híbrido baseado em *generalized autoregressive conditional heteroscedasticity* (GARCH) e filtro de Kalman com duas configurações. Na primeira o valor atípico era utilizado na previsão do passo seguinte e na segunda não. Os resultados mostraram que a previsão melhorou com a utilização do valor atípico e que a identificação desse valor se mostrava de forma mais pontual. Da análise dos gráficos apresentados no artigo verifica-se que a situação atípica é identificada no mesmo momento pelas duas abordagens utilizadas. Na segunda abordagem permanece mostrando valores atípicos ao longo de todo período em que ocorreram valores mais baixos, possivelmente indicando a duração do impacto do evento causador dessa redução. A identificação dos valores atípicos foi feita com base no quociente do desvio medido pelo desvio padrão, no entanto, neste caso o valor de corte utilizado foi o de 2,5 desvios padrão para amostras que incluíam os últimos 672 valores medidos. De modo a completar o entendimento verifica-se que nos modelos de Grubbs e Generalized Extreme Studentized Deviate (GESD), descritos neste trabalho, os valores de corte seriam respectivamente 3,8 e 1,64.

(PAN et al., 2013) efetuaram uma análise espaço temporal dos dados de vários sensores instalados em vias expressas e arteriais, considerando o impacto dos incidentes com base na velocidade, volume e ocupação. Concluíram que a inclusão da ocupação no modelo melhorou a determinação do impacto dos incidentes.

A fim de identificar os incidentes, (LI et al., 2015) elaboraram modelo em as séries temporais foram diferenciadas com base na média histórica diária e avaliaram a presença de valores atípicos na série diferenciada. Esta abordagem permite verificar variações atípicas, todavia pode ser menos eficiente quando avaliados dias atípicos como dias com chuvas fortes, feriados, etc.

A inclusão do fator espacial, baseada no tráfego de vias adjacentes, foi feita com um modelo de análise de vizinhança (*network neighborhood*). O modelo foi aplicado em vias

arteriais, considerando aspectos espaço temporais das ocorrências (SHI et al., 2018). Esta abordagem considera o impacto de incidentes de uma via em outra, todavia é dependente da existência de sensores nas diversas vias avaliadas como alternativas de rota da via onde ocorre o incidente.

Por fim, segundo (BURGHOUT; KOUTSOPOULOS; ANDREASSON, 2010), os simuladores de tráfego constituem ferramenta apropriada na reprodução e previsão dos efeitos de incidentes nas vias, uma vez que são capazes de incorporar a alteração na capacidade e nas características do fluxo e de redistribuir o tráfego pelas demais vias, considerando o impacto causado em toda a rede.

Estudos de previsão de acidentes com base nas condições de tráfego têm sido apresentados, e embora estejam fora do escopo deste estudo podem servir de base para a análise de incidentes. Neste tema temos muitos trabalhos que utilizaram os modelos hierárquicos de regressão inflados de zero (LORD; WASHINGTON; IVAN, 2005; LUKUSA; HING PHOA, 2020; MA; KOCKELMAN; DAMIEN, 2008).

a. Regressão multivariada inflada de zeros

O modelo de regressão multivariado inflacionado de zeros é baseado na combinação de uma distribuição de Bernoulli com uma distribuição de Poisson (ZIP) ou Poisson-Gama (ZIB), conforme existência ou não de superdispersão (FÁVERO; BELFIORE, 2017). A função resultante deste método de regressão possui dois processos geradores de zeros, sendo um dos chamados zeros amostrais, decorrentes da possibilidade de existência de zeros na amostra de contagem, e outro de zeros estruturais, relacionado com a distribuição binária que permite verificar a probabilidade da ocorrência ou não de determinada situação (FÁVERO; BELFIORE, 2017) e essa probabilidade pode ser obtida pelas expressões 4:

$$\begin{cases} p(Y_i = 0) = P_{logit_i} + (1 - P_{logit_i}) * e^{-\lambda_i} \\ p(Y_i = z) = (1 - P_{logit_i}) * \frac{e^{-\lambda_i} * \lambda_i^z}{z!}, \quad z = 1, 2, 3, \dots \end{cases} \quad (4)$$

$$P_{logit_i} = \frac{1}{1 + e^{-(\gamma + \delta_1 M_{1i} + \delta_2 M_{2i} + \dots + \delta_g M_{gi})}}$$

$$\lambda_i = e^{(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_g X_{gi})}$$

Onde Y é aproximadamente igual à Poisson Inflacionado de Zeros, M é a variável explicativa estrutural, X o vetor de variáveis explicativas da contagem, z identifica determinada contagem e α , β , γ e δ são os parâmetros a serem determinados pelo modelo.

b. Dados de ocorrências

Os dados de ocorrências devem ser utilizados na construção de métodos de identificação de incidentes e valores atípicos conforme proposto no algoritmo apresentado no capítulo 3.

Para o caso de São Paulo estão disponíveis os dados de registros das equipes de campo do órgão de trânsito do município (CET) e de um aplicativo de rede social que efetua roteamentos para seus usuários (Waze).

Presume-se ser necessário um relativo esforço no pré-processamento envolvendo a limpeza e organização dos dados tendo em vista a forma como são gerados e armazenados.

Após a etapa de pré-processamento os dados podem ser ordenados de acordo com o momento de início do fato, registrando-se a sua duração e, se disponível, sua ocupação na via.

Os bancos de dados de ocorrências assim ordenados podem ser cruzados com as séries temporais de características do tráfego ou com outras formas de monitoramento do fluxo, gerando uma base de estudo na qual seja possível efetuar análises cruzadas das características do fluxo com a existência de ocorrências.

i. Dados do órgão de trânsito

Podem ser obtidos diretamente do sistema de ocorrências do órgão de trânsito da cidade. No caso de São Paulo (CET) é possível obter os dados relativos à todas as ocorrências registradas pelos agentes de campo e pela Central de Operações.

As ocorrências são informadas em detalhes pelas equipes de campo quando da chegada ao local, indicando o local exato, que inclui o endereço e o sentido da via e a ocupação (calçada, número e posição das faixas de rolamento afetadas, acostamento, etc.). Essas informações são repassadas ao sistema de gestão da Central de Operações através de aplicativo de dados instalado em aparelhos celulares dos agentes, ou por meio de mensagens

de texto ou ainda por meio de comunicação por voz. O encerramento da ocupação da via também é informado pela equipe através do mesmo aplicativo.

As ocorrências são georreferenciadas com base nos endereços informados pelo agente de campo e armazenadas em bancos de dados relacionais, o que torna possível seu uso em tempo próximo ao real, através de integração com os bancos de dados do sistema de estimação aqui proposto.

No segundo semestre de 2018, foram registradas no sistema de dados da CET-SP 531.603 ocorrências, o que representa uma média de 88.600 ocorrências mensais sendo que nos dias úteis a média é de 3.500 registros. Destas, são relevantes ao estudo de incidentes cerca de 26% do total, sendo:

- 15% relativos à Incidentes: como veículos quebrados, obras e objetos deixados na via;
- 7% de falhas semafóricas: semáforos em falha crítica com impacto no fluxo;
- 3% de acidentes: com vítima, sem vítima e atropelamentos;
- 1% de outras ocorrências emergenciais: manifestações populares, alagamentos, tombamento de caminhões e queda de aeronave, dentre outras.

Outro aspecto importante com relação aos registros de ocorrências das equipes da CET é o fato de que é comum mais de um agente registrar a mesma ocorrência, sendo que alguns desses registros já são unificados pelo operador da Central de Operações durante o andamento da ocorrência.

Assim, para uma análise detalhada, é necessária a verificação dos registros efetuando a limpeza através de análise humana. Para confirmar este fato, em análise preliminar com os dados do segundo semestre de 2018 foi feito um filtro pelo número de registro da ocorrência tendo sido contatadas 6500 situações com mais de um registro no banco de dados.

Como as equipes da CET informam a quantidade de faixas afetadas por uma ocorrência, é possível incluir essa informação através da proporção de faixas de rolamento impactadas, uma vez que ocorrências que impactam mais faixas normalmente causam maior impacto ao fluxo da via.

ii. Dados de aplicativo de roteamento

A CET-SP, através de convênio específico, armazena a “imagem” do aplicativo de roteamento Waze a cada 30 minutos. Nessa imagem estão disponíveis os dados de

ocorrências que estavam ativas, sendo possível estabelecer horário em que algum usuário do aplicativo informou sua existência. Todavia, não é possível precisar o encerramento delas pois isso ocorre nos intervalos de gravação das imagens, sendo possível apenas identificar em que período de 30 minutos a ocorrência deixou de existir.

Em rápida análise dos dados foi possível constatar a existência de incidentes que não foram apontados nos registros de ocorrência da CET-SP. Assim, o uso dos dados de incidentes deste aplicativo, ainda que limitados pela forma como foram armazenados, podem auxiliar na elaboração dos modelos de detecção de incidentes.

Quanto ao posicionamento dessas ocorrências cabe indicar que sua geolocalização implica em imprecisões, já que a localização é atribuída à ocorrência quando da informação por um usuário o que pode ocorrer em local diferente devido à necessidade de manipulação do celular para efetuar-la. Assim, os registros muitas vezes ocorrem quando o usuário para numa espera semafórica ou estaciona o veículo para efetuar o registro.