

UNIVERSIDADE FEDERAL DO CEARÁ

PROGRAMA DE MESTRADO EM ENGENHARIA DE TRANSPORTES

**ANÁLISE COMPARATIVA DA APLICAÇÃO DE
MODELOS PARA IMPUTAÇÃO DO VOLUME MÉDIO
DIÁRIO DE SÉRIES HISTÓRICAS DE VOLUME DE
TRÁFEGO**

Antonia Fabiana Marques Almeida

**Dissertação submetida ao Programa de
Mestrado em Engenharia de Transportes
da Universidade Federal do Ceará, como
parte dos requisitos para a obtenção do
título de Mestre em Ciências (M. Sc.) em
Engenharia de Transportes**

ORIENTADORA: Prof.^a Dr.^a Maria Elisabeth Pinheiro Moreira

Fortaleza

2010

FICHA CATALOGRÁFICA

ALMEIDA, ANTONIA FABIANA MARQUES

Análise Comparativa da Aplicação de Modelos para Imputação do Volume Médio Diário de Séries Históricas de Volume de Tráfego, 2010.

93 fl., Dissertação (Mestrado em Engenharia de Transportes) – Programa de Mestrado em Engenharia de Transportes, Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2010.

- | | |
|--------------------------------|--------------------|
| 1. Análise de Séries Temporais | 2. Modelo ARIMA |
| 3. Monitoramento de Tráfego | 4. Modelos Simples |

CDD 388

REFERÊNCIA BIBLIOGRÁFICA

ALMEIDA, A. F. M. (2010). Análise Comparativa da Aplicação de Modelos para Imputação do Volume Médio Diário de Séries Históricas de Volume de Tráfego. Dissertação de Mestrado, Programa de Mestrado em Engenharia do Transportes, Universidade Federal do Ceará, Fortaleza, CE, 100 fl.

CESSÃO DE DIREITOS

NOME DO AUTOR: Antonia Fabiana Marques Almeida

TÍTULO DA DISSERTAÇÃO DE MESTRADO: Análise Comparativa da Aplicação de Modelos para Imputação do Volume Médio Diário de Séries Históricas de Volume de Tráfego. Fortaleza - Mestre / 2010.

É concedida à Universidade Federal do Ceará permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor.

Antonia Fabiana Marques Almeida

Av. Godofredo Maciel, 6257

60766-283 – Fortaleza/CE – Brasil

fabiana@det.ufc.br

ANÁLISE COMPARATIVA DA APLICAÇÃO DE MODELOS PARA IMPUTAÇÃO
DO VOLUME MÉDIO DIÁRIO DE SÉRIES HISTÓRICAS DE VOLUME DE
TRÁFEGO

Antonia Fabiana Marques Almeida

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE
MESTRADO EM ENGENHARIA DE TRANSPORTES DA UNIVERSIDADE
FEDERAL DO CEARÁ COMO PARTE DOS REQUISITOS NECESSÁRIOS À
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE
TRANSPORTES.

Aprovada por:

Prof.^a Maria Elisabeth Pinheiro Moreira, D. Sc.
(Orientadora)

Prof. Manoel Mendonça de Castro Neto, PhD.
(Examinador Interno)

Prof. Ricardo Almeida de Melo, D. Sc.
(Examinador Externo)

FORTALEZA, CE – BRASIL
SETEMBRO DE 2010

“Feliz aquele que ensina o que aprende e aprende o que ensina.”

Cora Coralina

AGRADECIMENTOS

Agradeço aos meus pais, Conceição e Veridiano, pelos ensinamentos de vida e formação como pessoa.

Ao meu marido, Marcos Timbó, pela ajuda, apoio, companheirismo e compreensão.

Ao meu irmão, Fabiano e minha família (tios, primos e avós).

Aos professores do agora IFCE, Marcelo Macedo, Hélio Henrique e Bartolomeu Cabral (*in memorian*).

À professora Elisabeth pela orientação, amizade e ensinamentos.

Aos professores membros da banca de defesa que muito auxiliaram com os comentários e considerações para melhoria do trabalho, Manoel e Ricardo.

Aos demais professores do PETRAN, em especial a professora Suelly e ao professor Waldemiro.

Aos amigos do mestrado Dmontier, Fátima, Filipe, Makey, e, em especial ao Breno, Bruno, Elton e Nauri.

À CAPES, pela bolsa que muito auxiliou nesta caminhada.

À amiga de longa data, Juliana Carla.

Ao professor Ary Ferreira, da UFC Cariri.

Aos demais amigos de Juazeiro do Norte, em especial Boris, Fernando e Lilian.

Ao DER/CE pela disponibilidade dos dados.

Resumo da Dissertação submetida ao PETRAN/UFC como parte dos requisitos para a obtenção do título de Mestre em Ciências (M.Sc.) em Engenharia de Transportes.

ANÁLISE COMPARATIVA DA APLICAÇÃO DE MODELOS PARA IMPUTAÇÃO
DO VOLUME MÉDIO DIÁRIO DE SÉRIES HISTÓRICAS DE VOLUME DE
TRÁFEGO

ANTONIA FABIANA MARQUES ALMEIDA

Setembro/2010

Orientadora: Maria Elisabeth Pinheiro Moreira

Para melhorias do sistema rodoviário, tanto no que se refere à infra-estrutura quanto à operação, é necessário a realização de estudos e planejamento, buscando a melhor utilização dos recursos existentes. Para tanto, faz-se o uso de uma importante medida de tráfego, o volume veicular. Os dados de tráfego são coletados por meio manuais ou eletrônicos, porém, ambos podem apresentar falhas e não coletar os dados em sua totalidade. No caso dos equipamentos eletrônicos de contagem, a coleta contínua pode formar uma série histórica, que, devido a não coleta, gera falhas ao longo da base de dados, as quais podem comprometer os estudos embasados nestas informações. Este trabalho busca, portanto, realizar análises de métodos empregados para estimação destes valores faltosos, buscando conhecer o modelo mais eficaz para a variável Volume Médio Diário dos dados obtidos pelos postos de contagem contínua instalados nas rodovias estaduais do Ceará. Os modelos de estimação aplicados neste trabalho são os modelos ARIMA de análise de séries temporais, e modelos simples, que apresentam aplicação menos complexa e processamento mais rápido, enquanto que o ARIMA demanda maior conhecimento específico do profissional que o utiliza. Assim, o método mais eficaz aqui considerado foi o que obteve menores erros após aplicação do modelo. Para estas aplicações foram selecionados quatro postos permanentes, de acordo com o percentual de dados válidos e sua localização, buscando a utilização de postos em pontos representativos do estado. O melhor modelo encontrado foi o ARIMA (1,0,1)⁷ (com erro médio de 1,816%), porém, um dos modelos simples, o MS2, obteve resultados próximos aos do ARIMA (erro médio 1,837%), e também pode ser considerado satisfatório para aplicação na imputação de valores faltosos.

Abstract of Thesis submitted to PETRAN/UFC as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.) in Transportation Engineering

COMPARATIVE ANALYSIS OF THE APPLICATION OF MODELS FOR THE
IMPUTATION OF AVERAGE DAILY VOLUME OF TRAFFIC VOLUME TIME
SERIES

ANTONIA FABIANA MARQUES ALMEIDA

September/2010

Advisor: Maria Elisabeth Pinheiro Moreira

In order to improve the road system, with regard to its infrastructure and operation, it is necessary to perform studies and planning, by seeking the best use of existing resources. Therefore an important traffic measure is used, i.e., vehicle volume. Traffic data is collected either manually or electronically; however both ways can fail and not collect all data. In the case of electronic counting equipment, the continuous data collection may form a time series, which produces failures in the database due to non-collection, which can compromise the studies based on this information. Therefore this work aims to perform analysis of methods used to estimate these missing values, by trying to know the most effective model for the Average Daily Volume variable of the data obtained by the continuous counting stations installed in the state highways of Ceará. The estimation models used in this work are the ARIMA models for time series analysis, and simple models, which present a less complex application and a faster processing, while the ARIMA requires more specific knowledge of the professional who uses it. The most effective method considered herein was the one that obtained smaller errors after the application of the models. Four permanent counting stations were selected for these applications, according to the percentage of valid data and its location, by seeking the use of stations in representative points of the state. The best model found was ARIMA (1,0,1)⁷ (with an average error of 1.816%), however one of the simplest models, MS2, produced results similar to those of ARIMA (an average error of 1.837%), and it can also be considered suitable for application in the allocation of missing values.

SUMÁRIO

CAPÍTULO 1	1
INTRODUÇÃO.....	1
1.1. CONTEXTUALIZAÇÃO	1
1.2. PROBLEMA DA PESQUISA.....	3
1.3. OBJETIVOS DA PESQUISA.....	5
1.3.1. Objetivo Geral.....	5
1.3.2. Objetivos Específicos.....	5
1.4. ESTRUTURA DO TRABALHO	6
CAPÍTULO 2	8
PROGRAMAS DE MONITORAMENTO DE TRÁFEGO	8
2.1. PROGRAMAS DE MONITORAMENTO DO VOLUME DE TRÁFEGO	8
2.2. QUALIDADE DOS DADOS DE TRÁFEGO	11
2.3. PROGRAMAS DE MONITORAMENTO DE TRÁFEGO NO BRASIL	13
2.3.1. Programa Nacional de Contagem do DNIT	13
2.3.2. Programa de Monitoramento do Volume de Tráfego nas Rodovias Estaduais do Ceará	16
2.4. FATORES DE EXPANSÃO	23
2.5. CONSIDERAÇÕES FINAIS	27
CAPÍTULO 3	28
MODELOS DE ESTIMAÇÃO DE DADOS	28
3.1. ANÁLISE DE SÉRIES TEMPORAIS – MODELO ARIMA	28
3.2. MODELOS SIMPLES APLICADOS NA INSERÇÃO DE DADOS FALTOSOS	40
3.2.1. Média Histórica.....	41
3.2.2. Média Histórica com Fator de Crescimento	41
3.2.3. Interpolação.....	42
3.3. MEDIDAS DE AVALIAÇÃO DAS TÉCNICAS DE IMPUTAÇÃO DE DADOS	44

3.3.1. Medidas Quantitativas.....	44
3.3.2. Medidas Qualitativas.....	48
3.4. ERROS CALCULADOS EM OUTROS ESTUDOS.....	49
3.4.1. Estudo de Zhong <i>et al.</i> (2006).....	49
3.4.2. Estudo de Zhong <i>et al.</i> (2003).....	50
3.5. CONSIDERAÇÕES FINAIS	52
CAPÍTULO 4	54
APLICAÇÃO E AVALIAÇÃO DOS MODELOS DE ESTIMAÇÃO DE DADOS FALTOSOS	54
4.1. ESCOLHA DOS POSTOS PERMANENTES E CARACTERIZAÇÃO DO VOLUME DE TRÁFEGO	54
4.2. IDENTIFICAÇÃO DOS MODELOS DE ESTIMAÇÃO ARIMA.....	59
4.3. ESCOLHA DOS MODELOS SIMPLES PARA ESTIMAÇÃO DE DADOS FALTOSOS	65
4.4. APLICAÇÃO E ANÁLISE DOS MODELOS DE ESTIMAÇÃO	67
4.5. CONSIDERAÇÕES FINAIS	72
CAPÍTULO 5	73
CONCLUSÕES E RECOMENDAÇÕES.....	73
5.1. CONCLUSÕES.....	73
5.2. RECOMENDAÇÕES	75
REFERÊNCIAS.....	76
ANEXO I	82
ANEXO II.....	84
ANEXO III	85

LISTA DE FIGURAS

Figura 2.1: Mapa com a distribuição espacial dos postos contadores da Malha Rodoviária Federal do Brasil (DNIT, 2008).....	15
Figura 2.2: Localização dos Postos de Contagem Permanente nas Rodovias Estaduais do Ceará (DERT, 2006)	17
Figura 2.3: Posto de Contagem Permanente Protegido por Abrigo de Concreto	18
Figura 2.4: Conjunto de Laços Indutivos Responsáveis pelas Contagens Contínuas Instalado no Pavimento.....	19
Figura 2.5: Percentual Médio de Dados Válidos, Inválidos e Faltosos de todos os Postos Permanentes Durante os anos de 1996 a 2005 (ENGEVIAS, 2009).	20
Figura 3.1: Série não-estacionária quanto ao nível e inclinação (BEZERRA, 2006). ..	30
Figura 3.2: Etapas iterativas da construção de modelos de previsão (Abordagem Box-Jenkins)	38
Figura 4.1: Percentual de Dados Válidos, Inválidos e Faltosos para todos os Postos de Contagem	55
Figura 4.2: Distribuição Espacial dos Postos Permanentes Seleccionados.....	56
Figura 4.3: Variação do Volume de Tráfego Entre os Meses do Ano	58
Figura 4.4: Evolução do Volume Diário de Veículos no Posto P002	59
Figura I.1: Evolução do Volume Diário de Veículos no Posto P003.....	82
Figura I.2: Evolução do Volume Diário de Veículos no Posto P011	82
Figura I.3: Evolução do Volume Diário de Veículos no Posto P012.....	83
Figura III.1: Autocorrelograma do Posto 003.....	85
Figura III.2: Autocorrelograma Parcial do Posto 003.....	85
Figura III.3: Autocorrelograma do Posto 011	86
Figura III.4: Autocorrelograma Parcial do Posto 011.....	86

Figura III.5: Autocorrelograma do Posto 012..... 87

Figura III.6: Autocorrelograma Parcial do Posto 012..... 87

LISTA DE TABELAS

Tabela 2.1: Descrição e Localização dos Postos Permanentes.....	17
Tabela 3.1: Resumo dos Modelos Simples Utilizados para Imputação de Dados de Tráfego.....	40
Tabela 3.2: Resumo de erros médios obtidos com a aplicação de modelos de estimação de dados em outros estudos.....	50
Tabela 3.3: Contadores Seleccionados e Dados Experimentais dos Diferentes Grupos..	51
Tabela 3.4: Erros Obtidos com a Aplicação dos Modelos de Imputação Avançados por Grupo.	51
Tabela 4.1: Volume Médio Diário, Desvio Padrão e Coeficiente de Variação dos Volumes Diários para os Postos Permanentes Seleccionados.....	58
Tabela 4.2: Testes de Normalidade e Estacionariedade para o Posto P002	60
Tabela 4.3: Parâmetros Para Seleção do Modelo ARIMA no Posto 002	62
Tabela 4.4: Tamanho das Amostras de Calibração dos Modelos ARIMA Para os Postos P003, P011 e P012	63
Tabela 4.5: Parâmetros Para Seleção do Modelo ARIMA no Posto 003	63
Tabela 4.6: Parâmetros Para Seleção do Modelo ARIMA no Posto 011	63
Tabela 4.7: Parâmetros Para Seleção do Modelo ARIMA no Posto 012	64
Tabela 4.8: Modelos ARIMA Seleccionados Para Cada Posto de Acordo com os Critérios AIC, BIC e MAPE	64
Tabela 4.9: Tamanho das Amostras de Cada Posto Para Aplicação dos Modelos.....	68
Tabela 4.10: Erros tipo MAPE obtidos com a aplicação dos modelos em relação ao percentual de dados perdidos do Posto 002	68
Tabela 4.11: Erros tipo MAPE obtidos com a aplicação de cada modelo com relação ao percentual de dados perdidos do Posto 003	69
Tabela 4.12: Erros tipo MAPE obtidos com a aplicação de cada modelo com relação ao percentual de dados perdidos do Posto 011	69

Tabela 4.13: Erros tipo MAPE obtidos com a aplicação de cada modelo com relação ao percentual de dados perdidos do Posto 012	70
Tabela 4.14: Resumo dos Erros dos Modelos Avaliados como Eficazes	71
Tabela II.1: Testes de Normalidade e Estacionariedade para o Posto 003	84
Tabela II.2: Testes de Normalidade e Estacionariedade para o Posto 011	84
Tabela II.3: Testes de Normalidade e Estacionariedade para o Posto 012	84

CAPÍTULO 1

INTRODUÇÃO

1.1. CONTEXTUALIZAÇÃO

O sistema de transporte é certamente um dos componentes fundamentais para o crescimento e desenvolvimento da economia de um país, visto que atende as necessidades de deslocamentos de bens e de pessoas. No Brasil, o modal rodoviário assume uma expressiva relevância em relação aos demais, respondendo por cerca de 96,2% do transporte de passageiros e a 61,8% do transporte de cargas (CNT, 2006).

No Estado do Ceará estes deslocamentos na sua maioria, também utilizam o sistema rodoviário, o que justifica a necessidade de um planejamento criterioso do setor, possibilitando uma aplicação racional dos recursos disponíveis. Portanto, para realizar estudos rodoviários de planejamento, projeto, manutenção, controle e operação, é necessário que os órgãos gestores sejam subsidiados com dados que representem a realidade das condições do tráfego, coletadas sob a forma de pesquisas de tráfego nas rodovias.

A importância desses dados se dá em virtude do processo de tomada de decisão que envolve frequentemente a alocação de recursos financeiros para a execução de melhorias na infra-estrutura das rodovias (AASHTO, 1992), possibilitando ao órgão rodoviário a aplicação racional dos recursos financeiros disponíveis.

Assim, o Departamento de Edificações e Rodovias do Ceará – DER/CE realizou estudo para implantação de um programa de monitoramento de tráfego, junto à ASTEF (Associação Técnica Engenheiro Paulo Frontin), concluído em 1994. Este programa tinha como objetivo principal o monitoramento contínuo do volume de tráfego nas rodovias estaduais do Ceará, permitindo a estimação do Volume Médio Diário Anual (VMDA) para todos os segmentos da malha rodoviária. Foi iniciada sua operação em 1996, buscando contribuir para a redução dos elevados custos de coleta e processamento de dados.

Segundo Zhong *et al.* (2006), os órgãos rodoviários gastam parcelas significativas em programas de monitoramento de tráfego. Em 2008, por exemplo, o Governo Estadual do Ceará realizou uma licitação para execução de pesquisa volumétrica e classificatória de tráfego na malha rodoviária estadual em somente 121 dos 264 postos existentes, totalizando um valor de aproximadamente R\$ 350.000,00 (ENGEVIAS, 2009). Ou seja, um total de equipamentos utilizados para a pesquisa, que corresponde a uma quantidade inferior que a metade dos existentes, foi gasto um valor significativo.

O programa de monitoramento de tráfego implantado pelo DER/CE é composto por postos de contagem contínua e postos de contagem de curta duração. Os postos de contagem contínua, segundo o DNIT (2006), são instalados em locais onde seja necessária uma série contínua de dados para a determinação de volumes horários, tendências dos volumes de tráfego, e principalmente permitir a determinação de fatores de expansão para corrigir contagens de curta duração. Estes postos, também denominados permanentes, funcionam 24 horas por dia, durante os 365 dias do ano.

Apesar da simplicidade na aplicação e obtenção dos fatores de expansão, motivos tais como o mau funcionamento dos equipamentos, vandalismo, falta de manutenção, falta de coleta, etc., podem fazer com que não se tenha disponível todos os dados de volume de tráfego ao longo do ano. Assim, pode não ser possível obter alguns fatores de expansão, ou podem ser calculados com base em dados incompletos, podendo comprometer os estudos rodoviários.

Mesmo que a base de dados de contagem contínua possua muitas falhas, o órgão rodoviário utiliza estas informações para obter o VMDA para expansão de contagens de curta duração, visto que são as únicas informações disponíveis. Portanto, não só a falta de dados, mas também a incerteza quanto à representatividade do VMDA, que porventura exista, podem gerar distorções nos estudos rodoviários.

Isto ocorre porque esta variável constitui-se do elemento básico à elaboração de planos diretores rodoviários, estudos de viabilidade técnico-econômica, projetos geométricos, projetos de pavimentos, programas de conservação, análise de segurança de tráfego, capacidade das vias, e outros.

1.2. PROBLEMA DA PESQUISA

Segundo Albright (1991a), desde os anos 30 há uma preocupação no tratamento e na análise de dados faltosos dos programas de monitoramento de tráfego, isso porque muitos órgãos rodoviários realizam a estimação de valores faltosos, também denominada de *imputation*. Essa técnica tornou-se popular entre os departamentos rodoviários nos Estados Unidos, pelo fato de que os altos percentuais de dias perdidos nos postos de contagem contínua compromete a determinação de parâmetros tais como o VMDA e os fatores de expansão, e devido às limitações financeiras e/ou tempo não é possível refazer uma nova contagem.

De acordo com a pesquisa desenvolvida por Albright (1991b), quando existiam falhas em equipamentos de contagem contínua (em 23 estados americanos) era comum a utilização de procedimentos de estimação de dados faltosos, e vários métodos eram usados para este propósito. Por exemplo, no Estado do Alabama, se menos de 6 horas de contagem fossem perdidas, esse dados eram estimados usando dados do ano anterior ou de outro mês. Entretanto se mais de 6 horas de contagem fossem perdidas, os dados do dia por completo eram anulados. No Estado de Delaware, as estimativas de dados faltosos eram obtidas da interpolação linear usando dados de meses antes e depois da falha. No estudo foi observado que na maioria destes departamentos rodoviários os métodos de imputação resumiam-se a aplicação de fatores simples obtidos de dados históricos e que não havia uma avaliação da precisão de tais estimativas.

A prática de *imputation* também é utilizada por órgãos rodoviários europeus, tais como Holanda, França e Reino Unido, e que segundo FHWA (1997), realizam a substituição de dados inválidos ou faltosos por dados históricos do mesmo local.

Além dos padrões históricos, também podem-se utilizar os padrões espaciais para a estimação de dados faltosos. Esse caso é mais aplicável em ambientes urbanos, onde são utilizados os sistemas *ITS - Intelligent Transportation Systems*, visto que em tais locais existe uma densa rede de detectores de tráfego. Smith *et al.* (2003) usaram, entre outros modelos, uma média ponderada entre um detector localizado a jusante, e outro a montante do detector com falha; conhecendo o padrão de distribuição do volume de tráfego entre as pistas, estimou-se o valor do volume para o *link* em falha.

Entretanto, segundo a AASHTO (1992), a estimação de dados faltosos não é recomendada, pois pode introduzir erros que não são quantificáveis, sendo sugerido aos órgãos rodoviários que se utilizam dessas práticas, que:

- Na base de dados devem ser identificados os valores perdidos que foram estimados. Isso assegura o Princípio Básico da Integridade dos Dados (ASTM, 1991), o qual afirma que as medidas de tráfego devem ser armazenadas sem quaisquer modificações ou ajustes;
- Documentar claramente os procedimentos usados e informar qual a magnitude da estimação no cálculo das estatísticas de tráfego obtidas;
- A técnica de estimação de valores faltosos deve ser temporária, visto que tais procedimentos não são tão importantes quanto à utilização e o tratamento dos dados reais coletados.

Porém, Sharma *et al.* (2003) afirmam que a estimação de dados faltosos não é proibitiva quando não se dispõe de outros dados, podendo-se utilizar esse procedimento. A grande vantagem do uso de valores históricos é a simplicidade, porém, Zhong *et al.* (2002) afirmam que os resultados obtidos por tais métodos normalmente são menos precisos que modelos mais sofisticados que utilizam, por exemplo, redes neurais, algoritmos genéticos e séries temporais.

A afirmação do trabalho de Zhong *et al.* (2002) foi obtida a partir de um estudo realizado com dados de contagem contínua de equipamentos instalados nas rodovias da Província de Alberta, Canadá, no qual foram comparados diversos modelos e técnicas de estimação de dados, sendo o melhor resultado obtido por modelos de séries temporais denominados ARIMA (Autorregressivo Integrado a Médias Móveis). Naquele estudo, utilizou-se 6 postos de contagem contínua que possuíam uma série histórica de 4 ou 5 anos sem falhas e os modelos propostos foram testados, retirando-se uma porção de dados existentes, simulando dados perdidos. Desses postos utilizaram-se dados de volume horário, e os modelos estimavam 168 volumes horários para uma semana de dados perdidos (7 dias x 24 horas), em diferentes meses do ano.

Entretanto, nas rodovias estaduais do Ceará, não se tem nenhum posto permanente com uma série histórica tão longa e sem qualquer tipo de falha. Para se ter idéia desta problemática, Gomes (2004) analisou a quantidade de dados disponíveis coletados por 12 postos de contagem permanente do DER, para o período compreendido entre os anos de 1996 e 2002, encontrando um percentual médio de dados válidos de 52%, de dados inválidos de 3%, e de dados faltosos ou perdidos de 45%. Vale lembrar também que a distribuição das falhas não ocorre em dias isolados, mas em alguns casos prolonga-se por semanas e/ou meses consecutivos, dificultando a criação de modelos precisos para a estimação de dados faltosos.

Devido a essas limitações, este trabalho pretende contribuir na investigação do uso de modelos de estimação de dados faltosos do volume de tráfego nas rodovias estaduais do Ceará, permitindo investigar a real eficácia destes para a imputação na base, levando em consideração a realidade local.

1.3. OBJETIVOS DA PESQUISA

1.3.1. Objetivo Geral

O objetivo geral deste trabalho é a verificação da eficácia de modelos de estimação de dados faltosos, com a aplicação de modelos simples e o ARIMA, buscando identificar qual o método que obtém resultados mais precisos para a série histórica de volume de tráfego dos equipamentos de contagem contínua instalados nas rodovias estaduais do Ceará.

1.3.2. Objetivos Específicos

Para que se possa atingir o objetivo geral anteriormente estabelecido, foi necessária a definição dos seguintes objetivos específicos:

- a) Apresentar os métodos utilizados para realizar estimação de dados de volume de tráfego;

- b) Caracterizar os volumes de tráfego dos postos de contagem contínua selecionados para aplicação dos modelos de estimação;
- c) Identificar o melhor modelo ARIMA através da etapa de calibração, para aplicação na etapa de validação, com o uso de amostras categorizadas com relação aos percentuais de dados faltosos e definir os modelos simples que serão aplicados;
- d) Realizar aplicação dos modelos de estimação em diferentes amostras com vários percentuais de dados faltosos e avaliar os erros de estimação obtidos em cada posto para cada tipo de modelo, buscando identificar assim qual apresenta maior eficácia.

1.4. ESTRUTURA DO TRABALHO

O trabalho de dissertação será apresentado em cinco capítulos. Este primeiro capítulo tratou da apresentação do trabalho a ser desenvolvido, com uma contextualização acerca do programa de monitoramento de tráfego e sua necessidade sistemática para o planejamento rodoviário e conseqüente redução de custos. Ainda neste capítulo é apresentada a problemática da falha de dados na série histórica obtida por meio dos postos de contagem permanente, instalados da malha rodoviária estadual do Ceará, justificando a importância do trabalho. Também é apresentado o objetivo geral e os específicos, visando o desenvolvimento do trabalho.

O segundo capítulo tem a finalidade de apresentar a estruturação de um programa de monitoramento de volume de tráfego, descrevendo seus objetivos, características e uma breve discussão sobre a forma de identificação e controle para manter a qualidade dos dados coletados. Também serão apresentados 2 programas de monitoramento, sendo o primeiro do órgão de jurisdição rodoviária federal, e o segundo do órgão de jurisdição estadual do Ceará, de onde foram coletados os dados que serão utilizados neste trabalho.

O terceiro capítulo trata da revisão bibliográfica realizada acerca do tema tratado por esta dissertação, que diz respeito aos modelos empregados na imputação de dados,

no caso o modelo ARIMA e os modelos simples. Também serão apresentados os métodos empregados em algumas localidades quando há a ocorrência de dados perdidos na série histórica de volume de tráfego, e os métodos de avaliação dos dados obtidos com a aplicação destes modelos.

No Capítulo 4 serão apresentados os dados dos postos de contagem contínua do Programa de Monitoramento das rodovias estaduais do Ceará, através de análises descritivas, e a aplicação dos modelos propostos, no caso o ARIMA e os modelos simples. Ainda neste capítulo também serão realizadas as análises dos dados obtidos através de comparações entre os modelos empregados.

No capítulo 5 serão realizadas as considerações finais do trabalho, através das conclusões obtidas com a aplicação dos modelos no capítulo anterior, bem como as recomendações para trabalhos complementares e futuros, buscando a melhoria da base de dados de volume de tráfego do programa de monitoramento aqui tratado.

CAPÍTULO 2

PROGRAMAS DE MONITORAMENTO DE TRÁFEGO

Neste capítulo pretende-se apresentar os programas de monitoramento de tráfego, com definições sobre variáveis e dados que fazem parte destes programas. Após isto, serão conceituados os Fatores de Expansão, que podem ser: semanal, mensal ou anual, cada um para uma finalidade, as quais são apresentadas. É abordada a qualidade dos dados de tráfego e, por fim, são apresentados 2 programas de monitoramento do Brasil: o federal, que abrange as rodovias federais brasileiras, e o estadual, que abrange as rodovias estaduais do Ceará, sendo deste programa os dados utilizados neste estudo.

2.1. PROGRAMAS DE MONITORAMENTO DO VOLUME DE TRÁFEGO

Para obter dados que representem a realidade das condições do tráfego e alcançar maiores êxitos nos estudos viários, muitos órgãos rodoviários criaram os programas de monitoramento de tráfego. Existe uma variedade de programas e essa diversidade reflete as diferenças entre os órgãos na aplicação de recursos financeiros para as rodovias, e automatização das contagens de tráfego (FHWA, 2001) envolvendo a coleta de muitos tipos de dados como volume, composição, velocidade e pesagem de veículos (MOHAMAD *et al.*, 1998).

Dentre os diversos tipos de dados coletados pelos programas de monitoramento de tráfego, conforme Oliveira (2004), o volume de tráfego é o mais básico de todos os parâmetros, e o mais ampla e freqüentemente usado como parte central de qualquer esforço na engenharia de tráfego. Esse parâmetro é definido como o número de veículos que passam por uma seção de uma via, em uma determinada faixa ou por sentido de circulação, durante um intervalo específico de tempo.

Para os estudos de planejamento de rodovias e estudos de tendências de crescimento do tráfego, o intervalo de tempo dos volumes de tráfego normalmente é o

dia (veículos/dia) sendo os parâmetros usados com mais frequência o VMD - Volume Médio Diário e o VMDA - Volume Médio Diário Anual (GOMES, 2004).

Com relação ao VMD, McShane e Roess (1990) afirmam que corresponde à média diária do volume de tráfego de uma determinada seção para um período menor que um ano, que pode ser medido para um período de seis meses, um mês, uma semana, ou menor período, como, por exemplo, um ou dois dias.

Já o VMDA corresponde à média dos 365 volumes diários registrados durante o ano em uma dada seção da via. De acordo com Mohamad *et al.* (1998), a determinação do verdadeiro valor desse parâmetro só é possível por meio de contagens contínuas, ou permanentes, efetuadas por equipamentos eletrônicos conhecidos como ATR's (*Automatic Traffic Recorders*), sendo normalmente os dados de volume registrados em períodos de uma hora e por faixa de tráfego.

Os postos de contagem são instalados em pontos onde se necessita de uma série contínua de dados, para determinação de distribuição temporal e estimativas de tendências. A sua distribuição deve ser tal que, os trechos onde estejam situados, sejam representativos de cada tipo de rodovia da malha, classificado de acordo com: função, situação geográfica, e volume de tráfego. As contagens nesses postos são realizadas durante as 24 horas dos 365 dias do ano, por meio de contadores automáticos. Os contadores devem ser locados em número suficiente e de maneira a cobrir toda a região em estudo, de modo que possa refletir as características do tráfego (GUIMARÃES, 2001).

A instalação de equipamentos de contagem em todas as seções da malha rodoviária é impraticável devido ao alto custo envolvido. Assim, um programa de monitoramento do volume de tráfego concentra-se, principalmente, na estimação do VMDA para todos os segmentos da malha rodoviária, adotando-se procedimentos baseados em amostras de contagem. Estas amostras são provenientes de dados dos postos de contagem de curta duração, implantados em pontos da malha viária onde se necessite, ou da estimação através dos dados obtidos por contagens de equipamentos de contagem contínua.

Segundo o FHWA (2001), o principal objetivo dos contadores contínuos é exatamente o desenvolvimento de fatores de expansão com o intuito de converter contagens de curta duração em estimativas médias anuais, eliminado o viés temporal no cálculo do VMDA. Contudo, DNIT (2006) lembra que a utilização de fatores de expansão sempre incorrerá em algum erro, pois jamais refletirá com exatidão os valores reais do posto, que seriam obtidos se fossem observados pelo período completo.

No entanto, a prática tem mostrado que este erro está dentro de limites aceitáveis na maioria dos casos, e que a economia de recursos é plenamente justificada. Naturalmente, as dimensões do erro, e a decisão dos seus limites de aceitabilidade, serão sempre função da precisão da pesquisa e dos critérios estabelecidos. O cálculo para a obtenção dos fatores de expansão será detalhado na seção 2.4.

As contagens de curta duração têm a função de assegurar a diversidade e cobertura geográfica do número máximo de pontos de malha rodoviária. Nesses postos normalmente são realizadas contagens manuais classificatórias com períodos que dependem do órgão rodoviário, sendo de 2 tipos:

a) **Postos Sazonais:** São postos temporários de contagem manual ou eletrônica, onde se registram os veículos que passam durante 24 horas para um período de 7 dias, sendo repetidas seguindo uma programação predeterminada, de três, quatro, seis ou doze vezes ao ano (ASTEF, 1993). Segundo o FHWA (2001), o principal objetivo deste tipo de posto é ajudar a identificar padrões de tráfego em vias específicas, como, por exemplo, em áreas turísticas, permitindo determinar o VMDA com maior precisão.

b) **Postos de Cobertura:** Nestes postos se realizam contagens uma vez no ano, durante 48 horas consecutivas em dois dias úteis da semana. O número de postos deve ser suficiente para completar a cobertura de todos os trechos da malha rodoviária. Não é necessário fazer contagem nesses postos todos os anos, à exceção daqueles necessários para comprovar variações previstas (DNIT, 2006). Entretanto, o FHWA (2001) afirma que as contagens de cobertura variam consideravelmente as durações e período de repetição de acordo com as necessidades, limitações orçamentárias e problemas enfrentados pelos órgãos rodoviários. Esse autor afirma que muitos estados americanos realizam

contagens de cobertura em períodos relativamente longos, de 2 a 7 dias, mas em apenas uma parte da malha rodoviária, a cada ano. Outros estados realizam uma contagem completa da malha rodoviária a cada ano, resultando em um grande número de postos de contagem, com duração de 24 ou de 48 horas.

2.2. QUALIDADE DOS DADOS DE TRÁFEGO

Após a coleta e armazenamento dos dados de campo os programas de monitoramento de tráfego devem assegurar aos usuários a utilização estrita de dados válidos, permitindo filtrar e eliminar as inconsistências da base de dados. O modo como os órgãos rodoviários identificam e controlam os erros dos dados de tráfego é um componente fundamental no programa de monitoramento.

Assim, esses programas devem manter registros de como os dados são manipulados, e que, cada manipulação tenha sido baseada em uma análise estatística. Não se devem excluir dados, ou simplesmente substituí-los, porque “eles parecem ser errados”. Ao invés disso, devem-se estabelecer procedimentos sistemáticos que permitam checar e identificar os dados inválidos, e, com isso, informar quando esses procedimentos de controle de qualidade são executados (FHWA, 2001).

Segundo Turner (2001), técnicas de controle de qualidade para dados de tráfego armazenados devem incluir a identificação e o tratamento de pelo menos três atributos:

2.2.1. Dado suspeito ou incorreto

Dados de valores ilógicos ou improváveis que não estão dentro de faixas esperadas ou regras estabelecidas. Nessa categoria de erro enquadram-se aqueles erros com valores de máximos e de mínimos extremos, como por exemplo, um valor de volume horário maior que a capacidade da via. Um critério usado para determinar a validade de um dado é compará-lo em relação a medidas históricas de um mesmo local, isso porque alguns padrões de características de tráfego regularmente se repetem.

2.2.2. Dados imprecisos

Dados de valores que são sistematicamente imprecisos (mas dentro da faixa de valores plausíveis), causados por erros de medidas do equipamento (ex. equipamento calibrado inadequadamente).

Um exemplo desse tipo de erro acontece quando o contador de tráfego registra consecutivamente o mesmo valor para um período longo de horas. Como as principais causas de inconsistências em uma base de dados de tráfego são ocasionadas por falhas, ou mau funcionamento dos equipamentos eletrônicos, tais erros podem provocar a repetição de valores sucessivamente (GOMES, 2004).

2.2.3. Dados perdidos

Dados de valores esperados que são perdidos por causa de mau funcionamento e/ou outros motivos. Esse atributo é o mais comum encontrado em quaisquer programas de monitoramento de tráfego, ocasionados, entre outros motivos, por falta de energia elétrica no local, por problemas nos equipamentos, falhas de comunicação, obras na via, etc.

Conforme Turner (2001), as características dos dados perdidos podem variar consideravelmente, dependendo do tipo de equipamento de monitoração do tráfego, dos controladores de campo e dos sistemas nas centrais de gerenciamento de tráfego, sendo importante não só identificar os dados perdidos na base de dados, mas também avaliar as possíveis causas da perda de dados. Isto requer análises de padrões nos dados perdidos.

Por conta disso, diversos procedimentos de qualidade de dados, desenvolvidos principalmente para aplicações de planejamento, prevêm orientação para estimar falhas em bases de dados de volume de tráfego. Esse processo de substituição de dados perdidos é também conhecido com imputação. (SHARMA, *et al.*, 2003; SMITH *et al.*, 2003), tornando-se uma prática freqüente nos programas de monitoramento de tráfego.

Albright (1993) adverte que, antes do uso indiscriminado da imputação de dados de tráfego, é necessária uma avaliação criteriosa do impacto dessa prática. Essa avaliação deve ser conduzida em função das características inerentes de cada local, incluindo o período e duração de dados perdidos, a técnica e o procedimento de imputação empregado, assim como a variabilidade local e especificamente do volume de tráfego.

No capítulo 3, serão discutidos alguns modelos de estimação (imputação) de valores faltosos de volume de tráfego nos programas de monitoramento.

2.3. PROGRAMAS DE MONITORAMENTO DE TRÁFEGO NO BRASIL

Para uma melhor compreensão acerca do estudo apresentado, se faz importante expor informações acerca dos Programas de Monitoramento de Tráfego da malha rodoviária estadual do Ceará e da federal brasileira, através de um histórico da sistemática que os órgãos gestores, com jurisdição sobre as malhas, têm realizado ao longo dos anos.

2.3.1. Programa Nacional de Contagem do DNIT

Através do Programa Nacional de Contagem de Tráfego (PNCT), o Departamento Nacional de Infra-Estrutura em Transportes (DNIT) vem nos últimos anos armazenando uma quantidade de informações de grande utilidade para os técnicos da área rodoviária. Apesar da desativação de alguns postos e criação de outros, o DNIT já dispõe de uma série histórica significativa de volume de tráfego em muitas rodovias nacionais.

A formação do PNCT ocorreu de forma lenta e gradativa, até chegar a sua atual constituição. O PNCT teve início em 1975, em Reunião de Técnicos de Trânsito, com a implantação do “Programa de Contagem Sistemática de Trânsito” em alguns estados brasileiros (Rio de Janeiro, São Paulo, Minas Gerais). Em 1976 este programa evoluiu para o “Plano Piloto de Contagem Sistemática de Trânsito”. Com o sucesso do plano

piloto e, verificada a eficiência dos aparelhos contadores, o DNIT expandiu o programa de contagens, empregando a mesma metodologia já testada e aprovada. Assim, em 1977 teve início o Plano Nacional de Contagem de Trânsito (PNCT) com 120 postos permanentes. Em 1989 o PNCT tinha implantado 235 postos permanentes, evoluindo para 266 postos em 1997, e para 285 postos em 1998 (DNIT, 2008).

O desenvolvimento do Plano Nacional de Contagem de Trânsito busca cobrir os trechos mais representativos da malha rodoviária de cada estado e, sem dúvida, de fundamental importância, pois seus resultados são subsídios básicos para os estudos de planejamento em geral, estudos econômicos, e os projetos rodoviários, essenciais ao estabelecimento de critérios para o cumprimento das seguintes finalidades:

- Planejar o sistema rodoviário;
- Programar necessidades e prioridades de melhorias no sistema rodoviário;
- Medir a demanda atual de serviços por via rodoviária;
- Estabelecer as tendências de tráfego no futuro;
- Determinar os volumes de viagens de forma a proporcionar justificativa econômica aos investimentos programados;
- Avaliar o volume existente de tráfego em relação ao sistema rodoviário atual;
- Estimar os benefícios dos usuários nas rodovias;
- Estabelecer uma classificação do sistema rodoviário;
- Justificar e planejar o policiamento;
- Estabelecer o veículo de projeto para fins de projeto geométrico;
- Projetar pavimento, obras de arte, seção transversal e outros elementos de rodovia;
- Realizar estudos de localização de postos de pesagem, socorro médico emergencial e etc.;

- Analisar a capacidade e estabelecer o nível de serviço;
- Realizar análise estatística de acidentes;
- Localizar e projetar instalações para a operação rodoviária.

A Figura 2.1 mostra a distribuição dos postos de contagem contínua ao longo das Rodovias Federais Brasileiras e observa-se que apesar da grande quantidade de postos existentes, atualmente todos estão desativados.

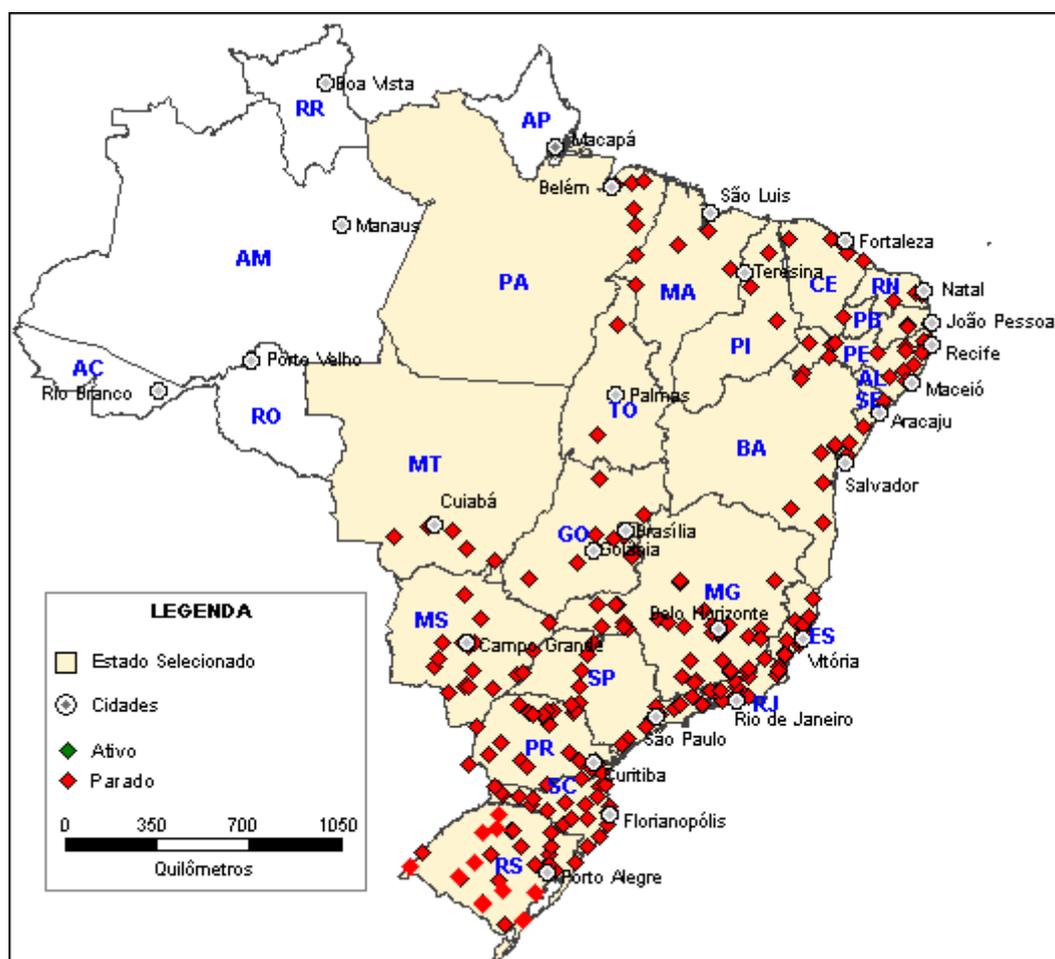


Figura 2.1: Mapa com a distribuição espacial dos postos contadores da Malha Rodoviária Federal do Brasil (DNIT, 2008)

Os equipamentos instalados nestes postos permitem realizar contagens volumétricas e classificatórias, por meio da implantação de um conjunto de instalações e aparelhos que exerçam distintas funções de captação, transmissão, detecção,

processamento e registro de volumes de volume de tráfego, compondo-se de quatro subconjuntos: abrigo, sensores de eixo, detectores e aparelho classificador. (DNIT, 2008).

Além desses postos de contagem contínua, na Rede Rodoviária Federal existia um total de 100 postos de monitoramento, entre planejados e em operação, que tinha como objetivo o levantamento de dados relativos ao volume de tráfego, às velocidades operacionais dos veículos, e às cargas por eixo para os veículos de carga e de transporte coletivo, em trechos rodoviários mais representativos, e funcionaram no período de 1999 a 2001. No *site* do DNIT estão disponíveis para *download* informações de 15 postos instalados em 7 estados.

2.3.2. Programa de Monitoramento do Volume de Tráfego nas Rodovias Estaduais do Ceará

A implantação do Programa de Monitoramento do Volume de Tráfego das Rodovias Estaduais do Ceará teve início com a execução de um Plano de Contagem do Tráfego no ano de 1994, resultante de um convênio firmado entre a Associação Técnico-Científica Eng. Paulo de Frontin - ASTEF / Universidade Federal do Ceará e o então Departamento de Edificações, Rodovias e Transportes – DERT, atual DER. (ASTEF, 1993).

A finalidade do programa era fornecer procedimentos eficientes para a obtenção de estimativas precisas do VMDA nos vários trechos rodoviários, baseados nas amostras de contagens. Esse programa de contagem de tráfego possuía 12 postos permanentes.

A Tabela 2.1 apresenta a descrição da localização destes postos. Para melhor visualização da distribuição espacial destes equipamentos, a Figura 2.2 ilustra a localização apresentada na Tabela 2.1 dos postos permanentes instalados nas rodovias estaduais do Ceará.

Tabela 2.1: Descrição e Localização dos Postos Permanentes.

POSTO	RODOVIA	SEGMENTO	
		INÍCIO	FIM
P001	CE-040	Entr. CE-453	Pindoretama
P002	CE-187	Entr. BR-222 (Tianguá)	Ubajara
P003	CE-060	Entr. BR-122 (B) / CE-368	Entr. CE-166/266 (Quixeramobim)
P004	CE-060	Entr. CE-284 (Umarizeiras)	Entr. BR-230 (A) (Várzea Alegre)
P005	CE-065	Entr. CE-251 (Jaçanaú)	Entr. CE-350 (Maranguape)
P006	CE-371	Entr. CE-138/265 (Morada Nova)	Entr. CE-266 (Roldão)
P007	CE-363	Entr. CE-060(B) (Mombaça)	Morada Nova
P008	CE-364	Entr. CE-313 (Campanário)	Entr. CE-232 (Moraújo)
P009	CE-176	Entr. CE-265 (Morro Redondo)	Entr. CE-266 (Tamboril)
P010	CE-293	Entr. CE-393 (B)	Entr. CE-153 (Missão Velha)
P011	CE-354	Entr. CE-243 (Varjota)	Entr. BR-222(A) (Umirim)
P012	CE-292	Araripe	Entr. CE-187/373 (Campos Sales)

Fonte: DERT (2006)

**Figura 2.2:** Localização dos Postos de Contagem Permanente nas Rodovias Estaduais do Ceará (DERT, 2006)

Conforme ENGEVIAS (2009), esses 12 postos permanentes iniciaram as contagens em 1996, operando até 2002. Após um período de inatividade, no ano de 2003, 8 postos voltaram ao funcionamento por mais 12 meses (Junho/2004 até Junho/2005). Os motivos que levaram à desativação dos 4 postos desativados foram:

- P001 (Iguape): o abrigo não oferecia segurança contra vandalismo, tendo sua porta arrombada, e por isso o equipamento eletrônico foi retirado do local;
- P007 (Mombaça), P009 (Tamboril) e P010 (Missão Velha): devido às péssimas condições do pavimento nestes locais, os laços indutivos foram danificados, sendo necessárias obras e preparos de elevados custos.

A coleta dos dados de tráfego nos postos permanentes era realizada através de equipamento eletrônico (modelo *Automatic Vehicle Classifier* – AVC100) fabricado pela *Pat Traffic Control Corp*. Estes equipamentos foram alocados de modo que o local pesquisado fosse representativo do trecho rodoviário, e foram instalados nas margens das rodovias em abrigos de concreto, protegidos de intempéries e vandalismo, como mostrado na Figura 2.3.



Figura 2.3: Posto de Contagem Permanente Protegido por Abrigo de Concreto

O processo de contagem desses postos era feito por um sensor do tipo laço indutivo, como ilustra a (Figura 2.4) instalado no pavimento, nos dois sentidos de tráfego da rodovia. Os valores registrados eram armazenados em uma memória em intervalos de uma hora, podendo ser em intervalo menor, desde que sejam configurados os parâmetros no *software* do equipamento. A capacidade de armazenamento dos dados no equipamento era de 20 dias.



Figura 2.4: Conjunto de Laços Indutivos Responsáveis pelas Contagens Contínuas Instalado no Pavimento.

Esse modelo de contador, tipo AVC, além de fazer a contagem volumétrica, permitia realizar a classificação automática de veículos, registrando até 7 tipos de categorias diferentes. Como o equipamento realizava a classificação de acordo com o comprimento do veículo, e não era possível distinguir um ônibus de um veículo de carga, caso o comprimento fosse igual. O conhecimento da composição do tráfego nos postos permanentes era conhecido por meio de pesquisas manuais. Essas contagens classificatórias manuais eram realizadas durante 24 horas em um período de 7 dias consecutivos a cada trimestre do ano, identificando 10 classes de veículos.

Os equipamentos também possuíam capacidade de transmitir os dados coletados por meio remoto, mas como esses contadores eram instalados em trechos rurais, não existia na época quaisquer tipo de linhas de transmissão. Assim, a coleta era realizada

por um técnico diretamente em cada posto, utilizando-se de um *notebook*, e caso o período da coleta ultrapassasse o período de armazenamento do equipamento, os primeiros dados registrados no último período eram perdidos (ASTEF, 2001).

Após a coleta em campo, os dados eram inseridos em um banco de dados que com o passar do tempo, tornou-se ineficiente e foi abandonado, pois não era robusto o bastante para trabalhar com uma grande quantidade de dados armazenados (GOMES, 2004).

Segundo ENGEVIAS (2009), foi criado um novo banco de dados contendo toda a série histórica de 1996 a 2005 das contagens permanentes, e utilizaram-se rotinas computacionais para filtrar e eliminar as inconsistências da base de dados. A Figura 2.5 ilustra os percentuais médios de dados válidos, inválidos, e faltosos de todos os postos permanentes, durante os anos de 1996 a 2005. Pode ser observado que o percentual médio de dados válidos é de 44%, de dados inválidos de 2%, e o de dados perdidos de 54%.

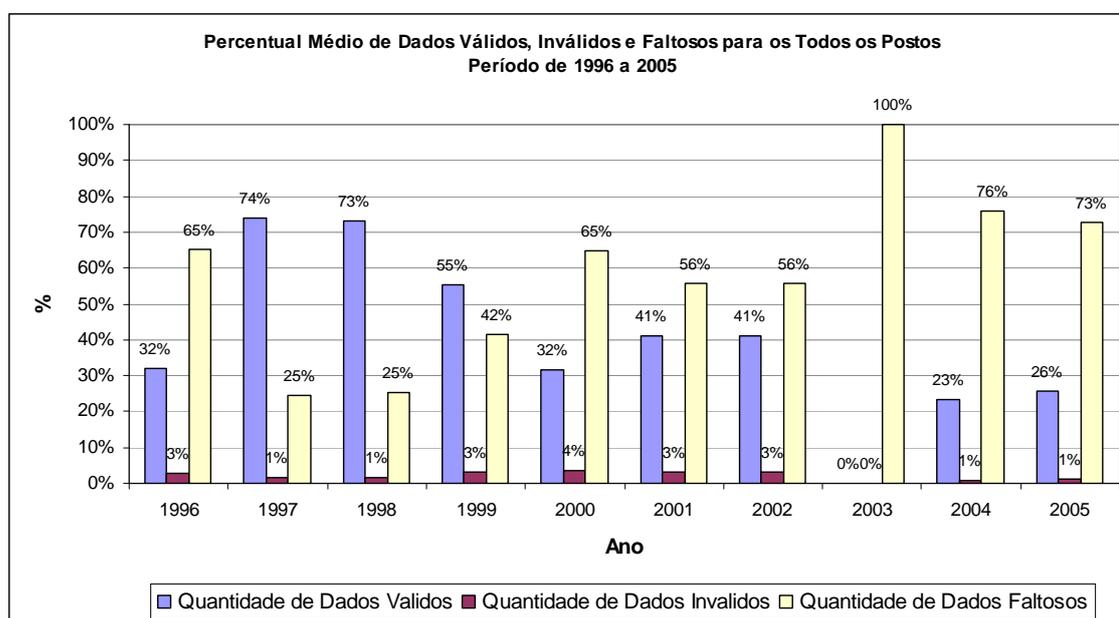


Figura 2.5: Percentual Médio de Dados Válidos, Inválidos e Faltosos de todos os Postos Permanentes Durante os anos de 1996 a 2005 (ENGEVIAS, 2009).

O alto percentual de dados perdidos apresentados na Figura 2.5, ficou mais evidente a partir do ano de 2000, resultando principalmente da falta de planejamento de

um plano de coleta dos dados armazenados nos contadores, aliado ao elevado custo de deslocamento de um técnico para a realização de tal atividade, e à ausência de um serviço de manutenção dos equipamentos eletrônicos, fazendo com que estes funcionassem de maneira precária até Junho/2005 e estando atualmente todos desativados. Além disso, quando ocorria uma falha no equipamento, essa se prolongava por vários dias consecutivos, e um dos principais motivos para isso é que não havia qualquer monitoramento sobre as falhas existentes nos postos permanentes. Também ocorria um grande intervalo de tempo entre as vistorias de manutenção/prevenção dos equipamentos (média de uma vez por mês).

Devido a esses problemas e outros fatores como tecnologia ultrapassada, elevado custo de manutenção, falta de peças de reposição, vandalismo, etc., o DER/CE utilizava-se somente das informações dos anos de 1998 e 1999 para a obtenção dos fatores de expansão, pois corresponde ao período de maior quantidade de dados coletados. A utilização dos dados de contagem desse período para a correção das atuais contagens de tráfego pode provocar erros, visto que tais dados possuem mais de 10 anos de defasagem, e que nesse período houve várias alterações físicas e operacionais na malha rodoviária, que podem ter provocado desvios e modificações no padrão de tráfego existente na década passada (ENGEVIAS, 2009).

Ainda de acordo com este autor, por consequência disso foi proposta a utilização dos equipamentos de fiscalização eletrônica, que realizam o controle de velocidade nas rodovias estaduais, para obtenção de fatores de expansão, visto que tais equipamentos também coletam informações dos volumes horários por sentido de tráfego, semelhantes aos contadores permanentes.

De acordo com ENGEVIAS (2009) foi realizada uma avaliação da qualidade e da quantidade de dados existentes dos equipamentos de fiscalização eletrônica, do período de 2002 a 2008, obtendo-se um percentual médio de dados válidos de 88 a 95%. Vale lembrar que estes equipamentos foram sendo implantados ao longo dos anos, portanto nem todos os locais possuem dados de toda a série histórica igual à obtida pelos postos permanentes. Além disso, alguns equipamentos funcionaram por algum período e após comprovada a redução de acidentes no local, eram transferidos para

outros trechos considerados pontos críticos em relação à segurança viária, deixando de realizar o monitoramento de tráfego do local inicial.

Segundo o FHWA (2001), a inclusão de outros tipos de equipamentos aos programas de monitoramento de tráfego é justificada, pois se deve tirar proveito da utilização de sofisticados equipamentos de coleta, que podem freqüentemente fornecer mais de um tipo de dado. Como exemplos podem ser usados sensores adicionais em balanças eletrônicas instaladas permanentemente para realizar classificação de veículos e contagem volumétrica, quando não são coletados dados de peso. Assim, uma balança do tipo WIN (*Weigh-In-Motion*) pode servir a três propósitos, reduzindo a necessidade da operação de outro dispositivo de coleta para dados adicionais neste local. Dessa maneira, permite ao órgão rodoviário a redução do número de equipamentos de coleta de dados operando continuamente e ao mesmo tempo aumentar a quantidade de dados disponíveis.

Albright (1993) também afirma que, apesar da diversidade existente entre os programas de monitoramento, nenhum programa de monitoramento de tráfego deve ser desenvolvido separadamente nos vários setores de um mesmo órgão. Isso porque os dados de tráfego são coletados para finalidades específicas, enquanto que, ocorrendo a integração desses dados em um banco comum, reduz-se a possibilidade de duplicação da coleta de dados em um mesmo local.

Apesar do grande número de equipamentos de fiscalização eletrônica instalados nas rodovias estaduais do Ceará, todos estão localizados em trechos rodoviários com características urbanas, sendo inadequados para a obtenção dos fatores de expansão, visto que:

- A localização de um posto de contagem deve ser cuidadosamente escolhida, de tal forma que o tráfego nesse ponto seja o representativo do trecho. Os limites do trecho deverão coincidir com pontos em que a composição e o volume de tráfego sofram variações significativas, sendo esses pontos extremos denominados pontos modificadores do tráfego, como por exemplo, cidades, vilas e entroncamentos com outras estradas (GEIPOT, 1970).

- Segundo o DNER (1976), os postos de contagem devem situar-se afastados das extremidades do trecho e de proximidades a qualquer gerador de tráfego, a fim de evitar distorções. Recomenda-se ainda que, uma vez escolhida a localização de um posto, todas as contagens futuras de tráfego no trecho em questão sejam feitas no mesmo local.

Os 12 postos permanentes atendem a estas condições, porém os equipamentos de fiscalização eletrônica, apesar de numerosos, possuem critérios e objetivos diferentes desses, de modo que deveriam agregar-se ao programa de monitoramento do volume de tráfego, gerando informações adicionais, mas não substituí-lo.

2.4. FATORES DE EXPANSÃO

Os fatores de expansão utilizados pelos órgãos rodoviários têm como principal objetivo o cálculo do VMDA de segmentos rodoviários onde foram realizadas as contagens de curta duração. Existe o método proposto por FHWA (2001) para a obtenção dos fatores, mas, antes de conhecê-lo, é importante saber quais os fatores de expansão podem ser obtidos dos postos de contagem contínua.

Segundo Moreira e Lima (1995), a hipótese envolvida neste processo de conversão baseia-se no fato de que trechos rodoviários com característica de tráfego semelhantes apresentam variações de volume de tráfego também semelhantes, dentro de uma mesma região, sujeitos, portanto a uma mesma variação sazonal, semanal e diária.

Então, a determinação do VMDA para uma contagem de tráfego de 24 horas, deve ser obtida de acordo com a Equação 2.1 (ASTEFL, 1993).

$$VMDA = VMD_d \cdot FES \cdot FEM \cdot FEA \quad (2.1)$$

onde,

VMD_d : Volume pesquisado na contagem de curta duração (24 horas);

FES: Fator de Expansão Semanal;

FEM: Fator de Expansão Mensal;

FEA: Fator de Expansão Anual.

2.4.1. Fator de Expansão Semanal (FES)

O Fator de Expansão Semanal (FES) é utilizado para transformar a contagem de curta duração obtida em um período ininterrupto de 24 horas no volume médio diário da semana, sendo calculado através da razão entre o Volume Médio Diário da Semana (VMD_{SEMANA}) e o Volume Médio Diário do Dia (VMD_{DIA}), como mostra a equação (2.2).

$$FES = \frac{VMD_{SEMANA}}{VMD_{DIA}} \quad (2.2)$$

onde,

2.4.2. Fator de Expansão Mensal (FEM)

O Fator de Expansão Mensal (FEM) é obtido através da razão entre o Volume Médio Diário do Mês ($VMD_{MÊS}$) pelo VMD_{SEMANA} , como pode ser verificado na equação (2.3).

$$FEM = \frac{VMD_{MÊS}}{VMD_{SEMANA}} \quad (2.3)$$

2.4.3. Fator de Expansão Anual (FEA)

O Fator de Expansão Anual (FEA) é obtido pela razão entre o VMDA do ano anterior e o $VMD_{MÊS}$, que corresponde ao Volume Médio Diário do mesmo mês da referente contagem, como pode ser verificado na equação (2.4).

$$FEA = \frac{VMD_{ANO}}{VMD_{MÊS}} \quad (2.4)$$

Vale ressaltar que a quantidade de fatores de expansão, ilustrados na Equação 2.1, pode ser reduzida, ou aumentada, em função das características dos contadores contínuos e do tempo da contagem de curta duração. Assim se, por exemplo, no posto permanente ao invés de contar veículos forem contados eixos (sensores pneumáticos), haverá a necessidade de adicionar um fator de ajuste para converter número de eixos em veículos. Entretanto, se a contagem for realizada durante uma semana não haverá necessidade do fator de expansão semanal.

Segundo o FHWA (2001), antes de utilizar os fatores de expansão alguns cuidados devem ser tomados, visto que algumas suposições estão envolvidas nesse processo de conversão:

- O agrupamento dos contadores permanentes em grupos de padrões similares dos fatores de ajuste mensal;
- A determinação dos fatores de ajuste médios para cada grupo;
- Atribuição a todas as seções do sistema de rodovias a estes grupos.

A principal dúvida na utilização dos fatores de expansão é a escolha do posto permanente do qual serão retirados os fatores para a correção da contagem de curta duração. Em revisões bibliográficas foram identificadas duas técnicas para a elaboração desse procedimento, tendo cada uma delas seus aspectos positivos e negativos. As técnicas discutidas brevemente a seguir são:

- Atribuição dos trechos rodoviários de acordo com características geográficas e classificação funcional;
- Aplicação do fator a uma mesma rodovia.

Segundo o FHWA (2001), na maioria dos estados americanos é usada uma combinação destas técnicas, e em alguns casos, a combinação dos métodos é provavelmente melhor, se comparado com o uso de somente uma técnica. Por exemplo, em seções rodoviárias onde existem os contadores contínuos nas proximidades, podem

ser aplicados fatores de contadores específicos nas contagens de curta duração destes trechos. Para os outros segmentos rodoviários, podem ser calculados e aplicados fatores médios de um grupo.

Fatores de grupos podem ser identificados usando uma técnica estatística de agrupamento, e em uma etapa seguinte pode-se usar o bom senso e o julgamento do profissional. Deste modo, novos ajustes podem ser feitos aos resultados de agrupamento para definir os fatores de grupo finais, de tal modo que possam ser facilmente aplicados. Também podem ser definidos inicialmente os grupos de acordo com o julgamento profissional e então modificados usando a técnica estatística. Muitos estados americanos utilizam os fatores de ajuste obtidos de um posto de contagem contínua mais próximo, ou da mesma rodovia, em postos de cobertura.

Este processo atribui os fatores de um único contador contínuo para todos os segmentos rodoviários dentro da área de influência daquele contador. O limite da zona de influência é definido em uma interseção que causa mudança significativa na natureza do volume de tráfego.

A contagem de curta duração em questão deve ser realizada próximo ao contador permanente; os fatores retirados do contador contínuo são aplicados, então, àquela contagem. A possibilidade de que a variação de tráfego do contador contínuo seja semelhante ao da contagem de curta duração é alta. As dificuldades na aplicação dessa técnica ocorrem somente quando as contagens de curta duração não estão próximas dos contadores permanentes. Em tal caso, os padrões de tráfego da contagem local podem ser diferentes dos encontrados no contador permanente. Logo, o método requer uma densa rede de contadores contínuos.

A retirada dos fatores de um único posto também oferece um inconveniente: caso ocorra um evento que provoque um tráfego atípico no posto permanente, tal efeito será transmitido a todos os trechos da área de influência, inclusive àqueles que não foram afetados pelo evento. Outro problema é que, se um posto permanente experimentar um período de longa parada, o cálculo dos fatores fica prejudicado.

Essas dificuldades podem ser superadas caso sejam usados fatores médios, retirados de um grupo de postos que possuam características e variação semelhantes do

tráfego. Assim, a disponibilidade de múltiplos contadores pode balancear a variabilidade do tráfego de um único local. Caso algum contador entre em falha, os fatores podem ser calculados utilizando os dados dos demais postos do grupo.

2.5. CONSIDERAÇÕES FINAIS

O presente capítulo fez uma breve apresentação sobre as principais características de um programa de monitoramento de tráfego rodoviário, identificando a finalidade de cada tipo de posto de coleta. Adicionalmente, discutiram-se as principais formas de identificação e controle para manter a qualidade dos dados coletados por postos de contagem contínua.

Foram também apresentados os tipos e as formas de cálculo dos fatores de expansão, produto básico de um programa de monitoramento de tráfego rodoviário, que permite a obtenção do VMDA para qualquer trecho da malha, a partir de contagens de curta duração.

Esse capítulo mostrou-se importante no conhecimento da estruturação e funcionamento de dois programas de monitoramento do volume de tráfego em rodovias, sendo o primeiro nas rodovias federais brasileiras e o segundo nas rodovias estaduais do Ceará. Esse último mereceu maior destaque, sendo detalhada a localização, a forma de coleta e tecnologia usada, bem como uma breve apresentação e discussão da atual situação do seu banco de dados.

Com o prosseguimento do trabalho, o capítulo a seguir trata da apresentação de alguns modelos utilizados por outros órgãos rodoviários para realizar a estimação de dados faltosos de volume de tráfego.

CAPÍTULO 3

MODELOS DE ESTIMAÇÃO DE DADOS

Este capítulo busca apresentar os métodos de estimação de dados, também conhecidos como métodos de imputação, a serem utilizados neste estudo. As primeiras técnicas de imputação de valores faltosos utilizadas eram métodos que envolviam substituição ou médias de dados, ou seja, técnicas simples. Depois surgiram técnicas como a interpolação temporal, a interpolação espacial, e outros, que serão apresentados neste capítulo. Como estes modelos apresentavam menor complexidade, e geravam dados menos precisos, as técnicas mais avançadas passaram a ser empregadas, conhecidas como análise de séries temporais, que podem ser os Algoritmos Genéticos, os modelos ARIMA, Redes Neurais, Funções Walsh, dentre outros.

Neste estudo é proposto o uso de modelos simples e dos modelos ARIMA, os quais serão apresentados de forma mais detalhada ao longo deste capítulo. No caso dos modelos simples, somente serão comentados os que foram selecionados para aplicação neste estudo. Também se buscou falar sobre as medidas de desempenho que auxiliam o conhecimento da precisão dos dados obtidos, bem como utilizá-los para comparação entre os dados obtidos, e estudos já realizados utilizando estes métodos.

3.1. ANÁLISE DE SÉRIES TEMPORAIS – MODELO ARIMA

De acordo com Ehlers (2007), uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. Os métodos para realizar a análise de séries temporais assumem que o conhecimento do passado e as experiências obtidas a partir desse conhecimento são úteis, pois, as experiências serão muito provavelmente repetidas no futuro (SHASHA, 1999 *apud* GARCIA, 2000). Para analisar estas séries, alguns modelos são propostos, os quais são diferentes entre si, tanto em sua formulação e quanto ao seu objetivo, e permitem identificar e descrever o comportamento de uma variável no tempo.

A característica mais importante de dados históricos é que as observações vizinhas são dependentes, e se está interessado em analisar e modelar estas dependências (EHLERS, 2007). Algumas características são particulares aos dados das séries temporais, como por exemplo:

- Observações correlacionadas são mais difíceis de analisar e requerem técnicas específicas;
- É necessário levar em conta a ordem temporal das observações;
- Fatores complicadores como a presença de tendências e variação sazonal, ou cíclica, podem ser difíceis de estimar ou remover;
- A seleção de modelos pode ser bastante complicada, e as ferramentas podem ser de difícil interpretação;
- É mais difícil lidar com observações perdidas e dados discrepantes devido à natureza sequencial.

A análise de séries temporais preocupa-se com a utilização de técnicas para a análise da dependência dos valores vizinhos. Isso requer o desenvolvimento de modelos estocásticos e dinâmicos para séries cronológicas e a utilização de tais modelos em áreas importantes de aplicação (BOX *et al.*, 1994). Existem situações em que o objetivo da análise de séries temporais pode ser a previsão de valores futuros, a relação com outras séries, ou ainda, a estrutura da própria série. Os principais objetivos da análise são (BEZERRA, 2006):

- Investigar o mecanismo gerador da série temporal;
- Descrever o comportamento da série, suas propriedades, através da verificação da existência de tendências, variações sazonais ou cíclicas, valores discrepantes, alterações estruturais, etc.;
- Predizer valores futuros da série com base nos valores passados, e estas previsões podem ocorrer a curto, médio, ou longo prazo;
- Procurar periodicidades relevantes nos dados;

- Explicar a variação de uma série através da análise da variação de outra série, considerando que esta interfira no desenvolvimento daquela, podendo auxiliar na explicação do mecanismo gerador da série, em que pode ser utilizada regressão múltipla, e pode-se estudar este problema, também, na teoria dos sistemas lineares (CORDEIRO, 2002);
- Controlar uma série temporal quando esta é uma medida da qualidade do processo de produção, havendo diferentes tipos de procedimentos de controle. Box e Jenkins são os responsáveis pelas mais sofisticadas estratégias e técnicas de controle (CORDEIRO, 2002).

Uma das suposições mais freqüentes que se faz a respeito de uma série temporal é a de que ela é estacionária. Ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Porém, a maioria das séries analisadas e encontradas apresenta alguma forma de não-estacionariedade, ou seja, apresentam alguma tendência. Estas tendências podem ser lineares, ou apresentarem alguma forma explosiva. Assim, uma série pode ser estacionária, não-estacionária homogênea, e não-estacionária explosiva (MORETTIN e TOLOI, 2006). A Figura 3.1 apresenta a forma de uma série não-estacionária.

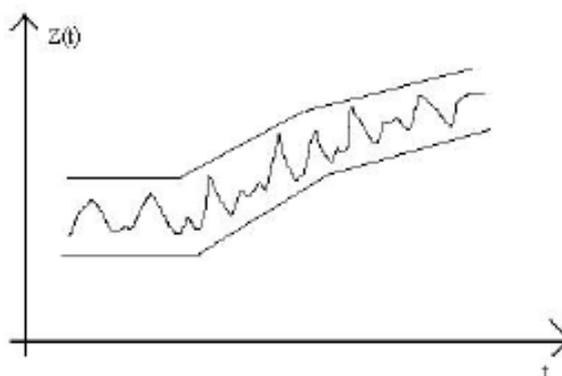


Figura 3.1: Série não-estacionária quanto ao nível e inclinação (BEZERRA, 2006).

Como a maioria dos procedimentos de análise estatística de séries temporais supõe que estas sejam estacionárias, quando não o são, há necessidade de transformação dos dados originais. A transformação mais comum consiste em tornar diferenças sucessivas da série original, até se obter uma série estacionária (MORETTIN e TOLOI, 2006). Porém, existem outras formas de tornar as séries estacionárias, mas aqui será

abordado somente o método de diferenciação, o qual é utilizado na aplicação de uma ou duas diferenças. A primeira diferença da série temporal $Z(t)$ é definida por:

$$\Delta Z(t) = Z(t) - Z(t-1) \quad (3.1)$$

onde:

$\Delta Z(t)$ é a diferença das séries $Z(t)$ e $Z(t-1)$;

$Z(t)$ é a série temporal no instante t ;

$Z(t-1)$ é a série temporal no instante anterior ($t-1$).

A segunda diferença é:

$$\Delta^2 Z(t) = \Delta[\Delta Z(t)] = \Delta[Z(t) - Z(t-1)] \quad (3.2)$$

Decompondo a equação da segunda diferença, se obtém:

$$\Delta^2 Z(t) = Z(t) - 2Z(t-1) + Z(t-2) \quad (3.3)$$

onde:

$\Delta^2 Z(t)$ e $\Delta[\Delta Z(t)]$ representam a segunda diferenciação aplicada;

$Z(t-2)$ é a série temporal no instante anterior ao ($t-1$), no caso o ($t-2$).

Caso outras diferenciações sejam aplicadas, como a terceira, quarta, e assim sucessivamente, ao final, se pode chegar à equação 3.4, que é a fórmula genérica para a diferenciação, independente da quantidade de ações que precisam ser executadas.

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)] \quad (3.4)$$

onde:

$\Delta^n Z(t)$ é a n diferenciação aplicada;

Tem-se, portanto, que a equação (3.4) é a Fórmula de Diferenciação. Estas diferenciações são aplicadas na série temporal em análise até que ela se torne estacionária. Assim, para os dados que não apresentam sazonalidade, a primeira diferença geralmente é suficiente para induzir a estacionariedade aproximada. Após a aplicação desta primeira diferenciação, é gerada uma nova série e, portanto, esta nova série pode requerer a aplicação de uma segunda diferenciação. Além de induzir a estacionariedade da série, a diferenciação pode ser aplicada em caráter exploratório. Por exemplo, caso existam observações discrepantes, uma representação gráfica pode ser suficiente para indicá-las.

Os modelos ARIMA, selecionados para aplicação neste estudo, são capazes de descrever tanto séries estacionárias quanto as séries não-estacionárias de forma satisfatória, desde que, no caso das séries não-estacionárias não apresentem comportamento explosivo, ou seja, sejam séries não-estacionárias homogêneas, não havendo necessidade de realizar a diferenciação.

Um importante exemplo de processo estacionário é o “ruído branco” (a_t), definido como uma sequência de variáveis aleatórias independente, identicamente distribuídas, e assim, pode-se afirmar que muitos processos podem ser construídos a partir do ruído branco. O termo “ruído branco” resulta do fato que em uma análise de frequência do modelo, pode-se mostrar que todas as frequências são iguais. Geralmente é assumido que o ruído branco tem média zero e variância σ_a^2 (BEZERRA, 2006).

Intuitivamente, um processo $X(t)$ é estacionário se ele se desenvolve no tempo, de modo que a escolha de uma origem dos tempos não seja importante (MORETTIN, 1999). Tecnicamente, há duas formas de estacionariedade: a fraca ampla (ou de segunda ordem), e a forte (ou estrita). Diniz *et al.* (1998) afirmam que uma série temporal é estacionária se o processo aleatório oscila em torno de um nível médio constante. Séries temporais sazonais com tendência linear, ou com tendência exponencial, são exemplos de séries temporais com comportamento não-estacionário.

Pode ser utilizado o teste de raiz unitária de Dickey-Fuller (DF) (DICKEY e FULLER, 1979) para verificação de estacionariedade, que testa a hipótese nula da existência de raiz unitária na série. Assim, quando a hipótese nula não pode ser

rejeitada, a série possui raiz unitária, o que significa que ela não é estacionária. Segundo Aêredes e Pereira (2008), o problema de autocorrelação dos resíduos faz com que o teste de raiz mais indicado seja o Dickey-Fuller Aumentado (ADF), pois este incorpora na equação de teste defasagens para a eliminação deste problema. Assim, se a hipótese nula for rejeitada neste caso, se pode concluir que a série não possui raiz unitária, ou seja, a série é estacionária. Caso a hipótese nula não possa ser rejeitada, não se deve aceitar de imediato que exista uma raiz unitária, e a equação de teste deve ser estimada através da inclusão, ou exclusão, dos termos intercepto e tendência, de acordo com suas significâncias estatísticas (ENDERS, 1995).

Ainda de acordo com Arêdes e Pereira (2008), o teste mais simples para analisar a estacionariedade em séries temporais é dado pela obtenção dos coeficientes ACF e PACF (Funções de Autocorrelação e Autocorrelação Parcial), que são as representações gráficas das defasagens. O correlograma traça um intervalo de confiança para os coeficientes no qual as estatísticas da ACF e PACF devem variar. Este intervalo é definido de acordo com a equação (3.5).

$$IC(95\%) = 0 \pm 1,96 \left(\frac{1}{\sqrt{n}} \right) \quad (3.5)$$

onde:

$IC(95\%)$ é o intervalo de confiança com 95% de significância estatística; e

n é o tamanho da amostra.

Gurajati (2000) afirma que coeficientes de autocorrelação e autocorrelação parcial fora deste intervalo, exceto para defasagem de 1 (um), são estatisticamente diferentes de zero, sugerindo a não-estacionariedade da série.

Muitas séries exibem uma variação periódica (diária, semanal, mensal, anual, etc.) a qual pode ser facilmente compreendida, medida e extraída dos dados, ofertando, assim, uma série dessazonalizada (CORDEIRO, 2002). Portanto, algumas séries temporais podem apresentar um comportamento que tende a se repetir a cada t período de tempo. Já a tendência pode ser definida como o conjunto de pares de valores observados, e o tempo para onde parece se dirigir esse conjunto de elementos

pertencente a uma série temporal (SILVA, 2005). De acordo com CORDEIRO (2002), são efeitos, em longo prazo, na média. Stevenson (1981) afirma que o termo “tendência” descreve um movimento suave dos dados de longo prazo, para cima ou para baixo, e podem estar relacionados com fatores, tais como: variações da população, modificações em preferências, etc.

De uma forma geral, uma série histórica pode exibir tendências de formas de crescimento e de decrescimento, e cada uma, com vários possíveis padrões, os quais estão descritos brevemente a seguir (EHLERS, 2007):

- Crescimento (ou decrescimento) Linear: é quando o aumento (ou redução) esperado a cada ano é o mesmo para todos os anos;
- Crescimento (ou decrescimento) Exponencial: é quando o aumento (ou redução) esperado a cada ano é dependente de um fator;
- Crescimento (ou decrescimento) amortecido: é quando o aumento (ou redução) esperado a cada ano é um percentual do aumento esperado do ano anterior.

Existem dois motivos que levam à necessidade de transformação dos dados originais (CORDEIRO, 2002): estabilizar a variância – se há uma tendência na série e a variância mostra crescimento com a média, então se indica uma transformação; tornar o efeito sazonal aditivo – se há uma tendência na série e o efeito sazonal aparenta crescer com a média, então é indicado fazer uma transformação.

Porém, de acordo com Morettin e Tolo (2006), Nelson (1976), Makridakis e Hibon (1979), Granger e Newbold (1976), as transformações não melhoram a qualidade da previsão, ou seja, os dados transformados têm pouco efeito na melhoria da previsão. Estes autores também afirmam que a heterocedasticidade não afeta na adequação da previsão, pois ela não implica em estimadores viesados, como no caso da regressão múltipla.

Ainda de acordo com estes autores, quando se tem um conjunto de dados que apresenta um padrão sazonal qualquer, é muito comum fazer um ajustamento sazonal dos dados, e depois usar um modelo não sazonal para se fazer a previsão. Já Plosser (1979) *apud* Morettin e Tolo (2006), analisa este problema para o caso dos modelos

ARIMA, e conclui que é preferível fazer a previsão usando diretamente o modelo sazonal, ao invés de ajustar sazonalmente a série, e depois utilizar um modelo não-sazonal. Outra razão para efetuar as transformações é obter uma distribuição para os dados mais simétrica e próxima da normal.

Os modelos para a análise de séries temporais variam dos mais simples aos mais sofisticados, em uma quantidade satisfatória para que se possa realizar a seleção da técnica mais apropriada à série que se vai analisar. Os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, controlados por leis probabilísticas (MORETTIN e TOLOI, 2006). A maioria dos métodos de previsão baseia-se na idéia de que as observações passadas contêm informações sobre o padrão de comportamento da série temporal. O propósito dos métodos é distinguir o padrão de qualquer ruído que possa estar contido nas observações, e então, usar esse padrão para prever valores futuros da série.

Na classe dos modelos paramétricos podem ser citados os modelos de erro (ou regressão), os modelos autoregressivos e de médias móveis (ARMA), os modelos autorregressivos integrados e de médias móveis (ARIMA), os modelos de memória longa (ARFIMA), os modelos estruturais, e os modelos não lineares. Já na classe dos modelos não-paramétricos, os mais utilizados são a função de autocovariância (ou autocorrelação) e sua Transformada de Fourier, o espectro. Neste trabalho serão utilizados os modelos ARIMA.

O ARIMA (*Autoregressive Integrated Moving Average*) é um modelo estatístico que utiliza variações e regressões de dados estatísticos com o fim de encontrar padrões para uma predição do futuro. Este modelo foi desenvolvido ao final dos anos 60. Box e Jenkins, em 1976, o sistematizaram (EHLERS, 2007) para determinar quanto do passado se deve utilizar para prosseguir na predição e os valores das séries.

O modelo ARIMA pode ser classificado como um modelo linear não-estacionário, de acordo com Box *et al.* (1994), e pode descrever processos que apresentam não-estacionariedade homogênea. São modelos populares devido a sua generalidade, podendo manipular quaisquer séries estacionárias, ou não. Existe uma quantidade considerável de *softwares* capazes de manipular os dados através do método ARIMA. A forma geral de um modelo ARIMA é apresentada na equação 3.6.

$$w_t = \mu + \frac{\Theta(B)}{\Phi(B)} a_t \quad (3.6)$$

onde:

t é o índice do tempo;

w_t é a d 'ésima diferença da variável de interesse z_t ;

$\Theta(B)$ é o operador de médias móveis: $\Theta(B) = (1 - \Theta_1 B^1 - \Theta_2 B^2 - \dots - \Theta_q B^q)$;

$\Phi(B)$ é o operador auto-regressivo: $\Phi(B) = (1 - \Phi_1 B^1 - \Phi_2 B^2 - \dots - \Phi_p B^p)$;

a_t é o “ruído branco” ou erro aleatório.

O ARIMA necessita que sejam identificados os coeficientes e o número de regressões que serão utilizados. Este modelo é bastante sensível à precisão com que se determinam seus coeficientes. É um modelo (p, d, q) , onde: ‘ p ’ se relaciona a Autorregressão, ‘ d ’ está relacionado à Integração ou Diferenciação, e ‘ q ’ a Média móvel. É um modelo dinâmico de séries de tempo, e utilizado para decidir as estimações futuras que venham a ser explicadas pelos dados do passado e por variações independentes. Para a construção dos modelos ARIMA, Box e Jenkins (1976) sugeriram as seguintes etapas iterativas.

a) Identificação → A identificação de um modelo ARIMA corresponde às determinações a seguir:

- nível de diferenciação (d), a partir do qual a série se torna estacionária;
- ordem (máxima) dos termos autorregressivos (p);
- ordem (máxima) dos termos médias-móveis (q).

Essas determinações são obtidas a partir do exame da função de autocorrelação e da função de autocorrelação parcial, que avaliam o padrão de dependência temporal da série. Para tanto, basta a comparação das funções de autocorrelação amostrais com os correspondentes paradigmas ou funções de autocorrelação teóricas.

b) Estimação → Uma vez determinada a ordem (p, d, q) do modelo, são estimados os seguintes parâmetros:

- μ , que é o nível do processo;
- os parâmetros auto-regressivos $\Phi_1, \Phi_2, \dots, \Phi_p$;
- os parâmetros de médias-móveis $\Theta_1, \Theta_2, \dots, \Theta_q$;
- σ^2 , a variância do ruído branco a_t .

c) Validação → A verificação da adequação do modelo é efetuada em duas dimensões inter-relacionadas, a saber:

- o exame do grau de ajustamento ou aderência do modelo – expresso em estatísticas como a variância do erro, como por exemplo, o erro quadrático médio (MSE – *Mean Square Error*), ou o percentual médio do erro absoluto (MAPE – *Mean Absolute Percentage Error*);
- o exame da aleatoriedade dos resíduos (erros) do modelo – expresso nas suas funções de autocorrelação.

d) Previsão → Após validado o modelo, pode ser construída uma função de previsão, que além de proporcionar as previsões mais verossímeis dentro do horizonte de planeamento especificado, proporciona também os limites superior e inferior do intervalo de confiança associado a um nível de probabilidade fornecido pelo analista.

Finalmente, é sugerido o emprego parcimonioso desses modelos, a sua utilização com um número mínimo de parâmetros. Assim, tendo em vista que a maioria das séries se torna estacionária após a segunda diferenciação, pode-se dizer que, usualmente tem-se: $p, d, q \leq 2$. Ou seja, no máximo a quantidade de auto-regressões, integrações e médias-móveis serão duas. A Figura 3.2 apresenta as etapas propostas para aplicação do modelo ARIMA.

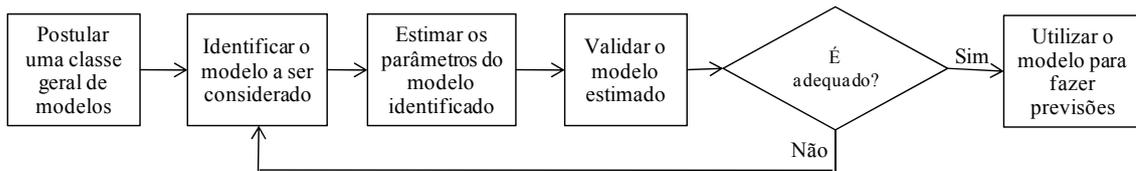


Figura 3.2: Etapas iterativas da construção de modelos de previsão (Abordagem Box-Jenkins)

Ou seja, a Figura 3.2 mostra de forma mais explicativa, como ocorre a aplicação do modelo Box-Jenkins, conforme segue:

- **Realizar diferenciação para obtenção de estacionaridade.** Para verificar se a série é estacionária, deve-se estudar o gráfico do autocorrelograma da série, pois, assim podem ser visualizadas características como a autocorrelação (FAC) e a autocorrelação parcial (FACP). O procedimento comum é a plotagem do correlograma de uma série y_t , e das sucessivas diferenças Δy , $\Delta^2 y$, e assim por diante, sendo observado a cada estágio, e se deve proceder desta forma até se obter a estacionaridade da série.
- **Identificar um modelo.** Após a diferenciação da série, buscando obter a estacionariedade da série, o autocorrelograma é examinado para definir a ordem apropriada dos componentes AR (Autorregressão) e MA (Média Móvel). O autocorrelograma de um processo MA é zero após dado ponto, enquanto que o de um processo AR declina de forma geométrica. Os autocorrelogramas dos processos ARMA mostram diferentes padrões.
- **Estimar o modelo.** Após a definição do modelo a ser utilizado, este deve ser aplicado para estimação. A estimação do modelo AR é direta, sendo estimado através da utilização do método de mínimos quadrados ordinários (MQO), buscando minimizar a soma dos erros quadrados (RSS): $RSS = \sum \varepsilon_t^2$. Já para os modelos MA, não é possível descrever a soma dos erros quadrados simplesmente como uma função dos valores observados e dos parâmetros. O que pode ser realizada é a descrição da matriz de covariância do erro da média-móvel e, se assumir normalidade, usa-se o método da Máxima Verossimilhança para Estimação (MVE).

No caso dos modelos ARMA, surge outro problema com a componente MA, porém, também pode ser utilizado o MVE.

- **Realizar diagnóstico.** Quando um modelo AR, MA ou ARMA é ajustado para uma série temporal, é aconselhável verificar se o modelo fornece uma descrição adequada dos dados. Existem dois critérios freqüentemente utilizados que refletem a qualidade do ajuste realizado e o número de parâmetros estimado. Um deles é o Critério de Informação de Akaike (AIC) (AKAIKE, 1973) e o outro é o Critério de Informação Bayesiano de Schwarz (SCHWARZ, 1978) (BIC). Assim, tem-se:

$$AIC(k) = n \log \sigma^2 + 2k \quad (3.7)$$

$$BIC(k) = n \log \sigma^2 + k \log n \quad (3.8)$$

onde:

k corresponde ao número total de parâmetros estimados;

n é o tamanho da amostra;

$\sigma^2 = \frac{RSS}{(n-p)}$, que representa o desvio padrão;

RSS é a soma dos resíduos quadrados.

Se vários modelos ARIMA estão sendo considerados, devem ser escolhidos os modelos que apresentem os menores valores para AIC e BIC. Estes dois critérios podem levar a diferentes conclusões, portanto, além destes, é importante verificar o padrão de correlação serial dos resíduos.

- **Previsão.** Supondo que tenha sido estimado um modelo com as observações, e se deseja prever y_{n+k} , que é considerada como a previsão de “ k ” períodos à frente dos valores observados.

Para facilitar, Box e Jenkins (1976) sugerem a adoção de um conjunto básico de funções para estimativa, que constitui o núcleo das expressões entre relações de medidas sobre elementos de séries temporais.

3.2. MODELOS SIMPLES APLICADOS NA INSERÇÃO DE DADOS FALTOSOS

Neste tópico serão apresentados alguns modelos simples utilizados por órgãos rodoviários em vários locais, aplicados por pesquisadores da área ou até para análise dos pesquisadores acerca dos modelos atualmente aplicados pelos órgãos. Zhong *et al.* (2005) reuniram em uma tabela alguns dos modelos simples, conforme pode ser visualizado na Tabela 3.1.

Tabela 3.1: Resumo dos Modelos Simples Utilizados para Imputação de Dados de Tráfego

Agência	Modelos de Imputação	Período Máximo de Imputação
Alabama	Volume horário do ano anterior ou de outro período do mês	6h
Alberta	Não imputa volumes horários faltosos, mas usam os dados históricos para estimar o Volume Médio Diário Mensal ($VMD_{MÊS}$)	Se menos de 1 semana de dados válidos não são coletados, não é calculado $VMD_{MÊS}$
Inglaterra (Grupo Central de Transporte)	Dados válidos da semana anterior coletados do mesmo equipamento	N/A
Delaware	Volumes horários da mesma hora do mês anterior e do mês posterior	N/A
França	Volume horário da mesma hora do mês anterior	N/A
Indiana	Dados do ano anterior	1 semana
Londres	Volume horário histórico das mesmas horas dos mesmos dias das semanas anteriores	N/A
Montana	Dados históricos da mesma localização	N/A
Holanda	Dados de outros equipamentos	N/A
Oklahoma	Dados dos mesmos dias das semanas do mês	9h
Saskatchewan (Manitoba)	Volumes horários da mesma hora do mesmo dia da semana do ano anterior	21 dias
Dakota do Sul	Volumes horários das mesmas horas do mesmo dia da semana dos anos anteriores	N/A
Vermont	Dados do mesmo dia do mesmo mês do ano anterior	N/A

Fonte: Adaptado de Zhong *et al.* (2005)

Através da verificação das informações da Tabela 3.1, pode-se notar que a maioria trabalha com estimação de dados horários de volume de tráfego. Também verifica-se que em algumas localidades é estabelecido um período máximo em que se pode realizar a imputação de dados (o maior período para efetuar a imputação foi de 21 dias). Os métodos utilizados baseiam-se em informações históricas levando em consideração dados do mesmo posto de períodos anteriores e/ou posteriores, bem como dados de outros postos.

A seguir são detalhados os principais modelos simples para estimação de dados faltosos do volume de tráfego, conforme revisão bibliográfica.

3.2.1. Média Histórica

Apresentado por Zhong *et al.* (2002), a média histórica assume que os volumes de tráfego tende a se tornarem estáveis após um período. Neste modelo, o valor perdido é estimado a partir da média dos dados históricos para a mesma hora, de anos anteriores existentes. É assumido neste modelo que não existem mudanças no volume de tráfego de ano para ano. A substituição dos valores perdidos é calculada de acordo com a Equação (3.9).

$$\text{Valor Perdido} = \frac{\sum_{i=1}^N \text{Valor Histórico}_i}{N} \quad (3.9)$$

em que,

N é a quantidade de observações.

3.2.2. Média Histórica com Fator de Crescimento

Este modelo também estudado por Zhong *et al.* (2002), acrescenta um Fator de Crescimento (FC) do tráfego da Equação 3.9. Os FC são usados como pesos no cálculo da média histórica, e é calculado usando a relação entre o VMDA do ano em que

ocorreu a falha e o VMDA dos anos de dados existentes. A substituição dos valores perdidos é obtida pela Equação 3.10.

$$\text{Valor Perdido} = \frac{\sum_{i=1}^N (\text{Valor Histórico}_i * FC_i)}{N} \quad (3.10)$$

em que,

N é o número de anos existentes;

$$FC_i = \frac{VMDA_{\text{Ano da Falha}}}{VMDA_{\text{Ano de Dados Existentes}}^i}$$

3.2.3. Interpolação

Este modelo foi apresentado por Chen *et al.* (2006), onde o algoritmo de interpolação temporal estima o valor faltoso através da substituição de uma média aritmética entre um valor anterior e posterior a falha, para o mesmo equipamento. Podem ser valores de uma semana anterior e de uma semana posterior, levando em consideração o mesmo dia da semana a que se refere a falha, conforme equação 3.10.

$$\text{Valor Perdido} = \frac{\left(\frac{VMS_i}{VMS_{(i-1)}}\right) * \text{valor}_{i-1} + \left(\frac{VMS_i}{VMS_{(i+1)}}\right) * \text{valor}_{i+1}}{2} \quad (3.10)$$

onde:

VMS_i é o Volume Médio Semanal da semana em que ocorreu a falha;

$VMS_{(i-1)}$ é o Volume Médio Semanal da semana anterior à falha;

$VMS_{(i+1)}$ é o Volume Médio Semanal da semana posterior à falha;

valor_{i-1} é o valor do volume do mesmo dia em que ocorreu à falha da semana anterior;

$valor_{i+1}$ é o valor do volume do mesmo dia em que ocorreu à falha da semana posterior.

Zhong *et al.* (2002), utilizaram o princípio de interpolação para estimar dados horários perdidos verificando o mês anterior e o mês posterior a falha, utilizando-se do fator mensal, como mostra a Equação 3.11.

$$Valor\ Perduto = \frac{\left(\frac{mf_i}{mf_{i-1}}\right) \times Valor_{i-1} + \left(\frac{mf_i}{mf_{i+1}}\right) \times Valor_{i+1}}{2} \quad (3.11)$$

em que,

mf_i é a média dos fatores mensais referente ao mês em que ocorreu a falha, calculados para os anos com dados existentes;

mf_{i-1} e mf_{i+1} são as médias dos fatores mensais para o mês anterior e posterior da falha respectivamente, em que são calculados para os anos existentes;

$Valor_{i-1}$ e $Valor_{i+1}$ são os volumes horários da mesma hora do mês antes e depois da falha respectivamente.

Os fatores mensais são obtidos de acordo com Equação 3.12.

$$Fator\ Mensal = \frac{Volume\ Médio\ Mensal}{Volume\ Médio\ Diário\ Anual} \quad (3.12)$$

O último modelo não incorpora o impacto do dia da semana, assim, Zhong *et al.* (2002) também propõem um modelo de Fatores Mensais e Dias da Semana, entretanto, $Valor_{i-1}$ e $Valor_{i+1}$ são volumes horários de uma mesma hora e dia da semana.

Os modelos citados anteriormente utilizam-se da interpolação temporal, determinado o valor faltoso a partir da média de um período antes e depois da falha para

o mesmo equipamento. Entretanto, há também órgãos que realizam a interpolação espacial, sendo essa técnica mais adequada para ambientes urbanos com densa rede de detectores veiculares que utilizam sistema ITS (*Intelligent Transportation System*). Chen *et al.* (2006) afirmam que conhecendo a correlação entre volumes de tráfego de locais geograficamente próximos, pode-se estimar o volume veicular faltoso para o mesmo instante, por meio da média entre os dados de volume de equipamentos instalados a jusante e a montante.

3.3. MEDIDAS DE AVALIAÇÃO DAS TÉCNICAS DE IMPUTAÇÃO DE DADOS

Como existem várias técnicas para imputação, ou previsão, de dados em séries temporais, há uma complexidade em conhecer qual das técnicas tem melhor eficácia. Para tanto, algumas formas de avaliação são propostas, através da análise de medidas quantitativas ou qualitativas. Ambos os tipos de medidas podem ser aplicados separadamente ou em conjunto, e algumas destas medidas serão apresentadas neste tópico.

3.3.1. Medidas Quantitativas

Medidas quantitativas são as que se pode mensurar, através de métodos e equações. Estas medidas fornecem um valor que, com base em outros ou no conhecimento do analista, servem de subsídio para avaliar os dados estudados. Assim, para se ter o conhecimento de métodos que podem ser empregados, os itens a seguir apresentam os mais utilizados, considerando que os dados utilizados neste estudo é o volume veicular.

- Medidas de Erro: o erro pode ser tratado como a diferença entre o valor atual e o observado, ou do valor imputado e o calculado. Uma formulação que descreve o cálculo do erro pode ser visualizada na Equação 3.13.

$$e_t = y_t - \bar{y}_t \quad (3.13)$$

em que,

e_t é o erro detectado para o instante t ;

y_t é o volume veicular atual ou observado, para o instante t ;

\bar{y}_t é o volume veicular calculado para o instante t .

Algumas medidas com base neste erro podem ser aplicadas para melhorar o cálculo do desempenho e obter valores mais precisos para conhecer a qualidade do modelo de estimação aplicado. Estas medidas podem ser vistas a seguir.

- MAE: o Erro Médio Absoluto (*Mean Absolute Error*) é uma média dos erros absolutos, e é uma medida comum de erro de previsão de análise de séries temporais. Pode ser calculado através da Equação 3.14.

$$MAE = \frac{\sum |y_t - \bar{y}_t|}{n} \quad (3.14)$$

em que,

n é o número de observações;

y_t é o volume veicular atual ou observado, para o instante t ;

\bar{y}_t é o volume veicular calculado para o instante t .

- MAPE: o Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error*) é calculado conforme a Equação 3.15, a seguir.

$$MAPE(\%) = \frac{1}{n} \sum \left| \frac{y_t - \bar{y}_t}{y_t} \right| \times 100 \quad (3.15)$$

y_t é o volume veicular atual ou observado, para o instante t ;

\bar{y}_t é o volume veicular calculado para o instante t .

- RMSE: a Raiz do Erro Quadrático Médio (*Root Mean Squared Error*) é um peso médio do erro que aplica um peso maior para grandes erros, do que para erros pequenos. Esta é uma medida clássica de desempenho muito utilizada no desenvolvimento e análise do modelo. Pode ser obtida por meio da aplicação da Equação 3.16.

$$RMSE = \sqrt{\frac{\sum (y_t - \bar{y}_t)^2}{n}} \quad (3.16)$$

- SDE: o Desvio Padrão dos Erros (*Standard Deviation of Errors*) é uma medida de quanta variância existe em um conjunto de erros. Um desvio padrão mais baixo significa que os erros estão fortemente agrupados ao redor do valor médio. O SDE pode ser obtido através da aplicação da Equação 3.17.

$$SDE = \sqrt{\text{var}(y_t - \bar{y}_t)} \quad (3.17)$$

y_t é o volume veicular atual ou observado, para o instante t ;

\bar{y}_t é o volume veicular calculado para o instante t .

- MinAE: o Erro Mínimo Absoluto (*Minimum Absolute Error*) é o menor valor de erro absoluto. Pode ser obtido através da Equação 3.18.

$$MinAE = \min(y_t - \bar{y}_t) \quad (3.18)$$

- MaxAE: o Erro Absoluto Máximo (*Maximum Absolute Error*) é o maior valor de erro absoluto. Pode ser obtido através da Equação 3.19.

$$MaxAE = \max(y_t - \bar{y}_t) \quad (3.19)$$

- MSE: Erro Quadrático Médio (*Mean Square Error*) é uma medida de quanto os dados estão próximos da linha de ajuste do erro. Para tanto, basta somar todas as diferenças tomadas no eixo y entre ambos, aplicando a Equação 3.20.

$$MSE = \frac{\sum (y_t - \bar{y}_t)^2}{n} \quad (3.20)$$

- U de Theil (THEIL, 1966): este método é conhecido como coeficiente de desigualdade, no qual se pode ter uma idéia das discrepâncias cometidas pela previsão, atuando como um parâmetro de avaliação (quanto mais próximo de zero é o valor obtido, melhor é o desempenho do modelo). Pode ser obtido através da aplicação da Equação 3.21.

$$U^2 = \frac{\sum_{t=1}^{n-1} \left(\frac{\bar{y}_{t+1} - y_{t+1}}{y_t} \right)^2}{\sum_{t=1}^{n-1} \left(\frac{y_{t+1} - y_t}{y_t} \right)^2} \quad (3.21)$$

onde:

y_{t+1} é valor observado no instante (t+1);

\bar{y}_{t+1} é valor calculado no instante (t+1).

y_t é o volume veicular atual ou observado, para o instante t ;

n é o tamanho da amostra.

- APE: Erros Percentuais Absolutos (*Absolute Percentage Errors*). Pode ser obtido através da aplicação da Equação 3.22, a seguir apresentada.

$$APE = \frac{|y_t - \bar{y}_t|}{y_t} \quad (3.22)$$

y_t é o volume veicular atual ou observado, para o instante t ;

\bar{y}_t é o volume veicular calculado para o instante t .

Estas medidas também foram utilizadas em outros estudos para verificar a eficácia dos modelos aplicados. Ainda há outro método (apresentado no Capítulo 4) que foi escolhido para ser também utilizado neste estudo, por permitir a avaliação em

percentuais dos erros calculados após a estimação dos dados através do uso dos modelos simples e ARIMA.

3.3.2. Medidas Qualitativas

As medidas qualitativas avaliam parâmetros que, a princípio, não podem ser mensurados, e envolve juízo, buscando excluir arbítrio e subjetividade. Para o caso de análise de modelos de inserção de dados, a seguir são descritos exemplos de alguns métodos que podem ser utilizados para avaliação.

- Dados de entrada requeridos: diferentes técnicas de imputação de dados requerem diferentes tipos de dados de entrada. Esta medida é uma avaliação da disponibilidade dos dados de entrada requeridos. Se uma técnica requer muitos tipos de dados diferentes, a probabilidade de todos estes dados estarem disponíveis em uma base consistente pode ser muito baixa (SMITH *et al.*, 2003). Assim, busca-se avaliar a quantidade de dados requeridos, a correlação entre as variáveis que o modelo requer, e a disponibilidade de todos os dados de entrada solicitados.
- Complexidade: algumas técnicas de imputação são bastante simples e de fácil entendimento, enquanto outras requerem grande conhecimento. Assim, enquanto alguns dos modelos mais complexos podem prover boas estimativas, eles podem ser de difícil utilização e podem ser impraticáveis para utilização por parte de profissionais e pesquisadores da área (SMITH *et al.*, 2003). Assim, deve-se verificar os quesitos que abrangem a dificuldade de aplicação do modelo, o quanto se deve ter conhecimento acerca do método a ser implementado, e a precisão dos dados obtidos após a aplicação do modelo, visto que, modelos mais simples podem obter resultados similares se comparado a modelos mais complexos, por exemplo.
- Velocidade computacional: Técnicas para a estimativa de dados de tráfego podem ser amplamente classificadas em duas categorias: heurísticas ou estatísticas. Quando estas técnicas são aplicadas em ambiente computacional, levam diferentes medidas de tempo para que os cálculos sejam concluídos e, assim, fornecer o resultado final. Para avaliar este parâmetro, o usuário pode optar como melhor

método o que dure mais ou menos tempo para ser calculado. O ideal seria a avaliação deste critério com algum outro, dada a sua particularidade, dependendo dos objetivos da análise e dos dados utilizados (CHEN *et al.*, 2006).

3.4. ERROS CALCULADOS EM OUTROS ESTUDOS

Aqui serão apresentados resultados obtidos da aplicação de modelos de estimação de dados para imputação em outros estudos, buscando estabelecer faixas de erro aceitáveis para embasar a análise dos erros deste estudo.

3.4.1. Estudo de Zhong *et al.* (2006)

Este estudo avaliou os modelos usados em Saskatchewan, Dakota do Sul, Delaware, França e Londres, utilizando-os para estimação de dados em uma base do Canadá. No modelo usado em Saskatchewan, os dados faltosos são imputados com o uso dos dados do ano anterior e correspondendo ao mesmo dia da mesma semana. Em Dakota do Sul os valores faltosos são estimados com a média dos dados dos mesmos períodos dos 3 anos anteriores. Já Delaware estima os dados faltosos baseados na média de dados do mesmo período dos meses adjacentes.

A França usa os dados dos meses anteriores para estimar valores faltosos de volumes horários de vários dias (exemplo: quartas e sábados) e de diferentes estações (exemplo: verão e outono). Londres usa um modelo de média móvel ponderada exponencialmente. Este modelo envolve a validação de dados de novas contagens de tráfego com referência em dados antigos do mesmo equipamento, coletados nas semanas anteriores no mesmo período. A Equação 3.23 descreve o modelo utilizado por Londres.

$$\hat{x}_{t,s} = (1-\theta)x_{t-1,s} + (1-\theta)\theta x_{t-2,s} + (1-\theta)\theta^2 x_{t-3,s} + \dots + (1-\theta)\theta^{n-1} x_{t-n,s} \quad (3.23)$$

Na Tabela 3.2 estão apresentados os dados dos erros obtidos no estudo de Zhong *et al.* (2006), através da utilização dos modelos das cidades citadas nesta seção, e do modelo ARIMA, sendo o modelo (p,d,q)^s utilizado o (1,1,0)¹².

Tabela 3.2: Resumo de erros médios obtidos com a aplicação de modelos de estimação de dados em outros estudos

Hora	Média dos Erros de Predição					
	Saskatchewan	Dakota do Sul	Delaware	França	Londres	ARIMA
07-08	10,71	13,11	7,65	10,75	6,99	4,86
08-09	7,45	9,99	5,42	9,02	4,41	4,67
09-10	5,74	6,37	3,18	5,17	3,27	3,03
10-11	4,96	5,32	4,65	6,04	3,89	2,87
11-12	5,75	2,94	5,68	9,00	4,47	3,71
12-13	6,29	4,07	5,42	8,28	4,28	3,43
13-14	5,16	7,59	4,91	7,17	4,18	3,31
14-15	4,93	7,48	3,39	4,71	3,04	4,47
15-16	5,38	3,96	2,71	3,78	2,59	2,38
16-17	5,98	2,96	3,29	4,70	3,31	4,46
17-18	6,34	4,24	4,17	5,47	3,29	4,03
18-19	6,31	10,26	4,32	6,25	3,98	5,74
Média Total	6,25	6,52	4,57	6,70	3,98	3,91

Adaptado de Zhong *et al.* (2006)

Pode-se observar que a média total dos erros de estimação de dados de volume no estudo foram inferiores a 7, e para os casos dos modelos mais complexos, o modelo Londres e o modelo ARIMA, foram inferiores a 4. Os autores avaliam como mais eficaz, dentre os modelos analisados, o ARIMA, por ter obtido uma média total inferior aos demais. Porém, não descarta o uso dos demais modelos, pois, por exemplo, no caso do modelo de estimação utilizado em Delaware, a diferença entre a média total obtida por esta aplicação e a do modelo ARIMA foi considerada baixa.

3.4.2. Estudo de Zhong *et al.* (2003)

Neste estudo são utilizados os dados do órgão rodoviário da cidade de Alberta, que utiliza 350 contadores veiculares. Foi proposto um método de agrupamento por Sharma e Werner (1981), que classificam os contadores de tráfego em grupos. Depois de realizar uma análise do padrão dos grupos entre os anos 1996 e 2000, foram obtidos cinco grupos. Estes grupos foram classificados como: de tráfego geral, de tráfego local, rural, turística de verão e turística de inverno.

Seis contadores permanentes foram selecionados de quatro dos cinco grupos acima mencionados: dois do grupo de tráfego geral, dois do grupo de tráfego local, um do grupo rural, e um do grupo turístico. Devido aos dados insuficientes no grupo turístico de inverno, nenhum contador deste grupo foi selecionado. A Tabela 3.3 mostra os contadores selecionados dos diferentes grupos, VMDA, e os períodos de calibração e análise.

Tabela 3.3: Contadores Selecionados e Dados Experimentais dos Diferentes Grupos

Nome grupo	Grupo de padrão de viagem	Classe da via	VMDA	Período de Calibração	Período de Análise
CM1	Tráfego Geral	Arterial principal	41575	1996-1999	2000
CM2	Tráfego Geral	Coletora	4042	1996-1999	2000
RC1	Tráfego Local	Coletora	3580	1996-1999	2000
RC2	Tráfego Local	Coletora	3905	1996-1999	2000
RLD	Rural	Arterial	13627	1996-1999	2000
SR	Turística de Verão	Coletora	2002	1996-1998	2000

Adaptado de Zhong *et al.* (2003)

Para imputação dos dados perdidos foram aplicados os modelos tradicionais, que foram os mesmos utilizados no estudo de Zhong *et al.* (2006), apresentados na seção 3.4.1, e os modelos avançados levaram em consideração regressões baseadas em algoritmos e modelos de redes neurais. Os erros obtidos com a aplicação dos modelos tradicionais estão apresentados na Tabela 3.2, e os erros obtidos com a aplicação dos modelos avançados estão na Tabela 3.4 a seguir.

Tabela 3.4: Erros Obtidos com a Aplicação dos Modelos de Imputação Avançados por Grupo.

Hora	Regressão baseada em Algoritmos						Redes Neurais					
	CM1	CM2	RC1	RC2	RLD	SR	CM1	CM2	RC1	RC2	RLD	SR
07-08	0,25	2,71	2,51	1,66	1,33	6,08	2,14	24,59	10,51	5,81	5,95	17,12
08-09	0,59	2,34	2,20	3,26	0,90	1,07	2,19	13,52	4,67	2,91	3,39	10,72
09-10	1,66	2,50	1,98	2,03	1,34	0,76	3,03	6,55	3,34	8,38	2,95	12,10
10-11	0,38	1,66	0,83	1,93	2,08	3,78	6,71	6,97	5,07	4,10	9,30	9,29
11-12	0,56	0,82	1,19	1,29	2,12	1,43	2,13	4,86	3,57	4,49	7,60	12,47
12-13	0,81	2,76	1,10	2,76	2,09	1,70	2,90	12,72	7,79	5,10	7,81	6,09
13-14	0,59	1,71	1,64	2,83	1,39	1,34	2,23	11,56	6,25	7,29	12,18	6,66
14-15	0,89	2,50	3,92	1,82	1,43	1,81	4,69	8,10	5,13	5,84	15,01	7,73
15-16	0,68	3,83	0,49	0,82	1,22	1,83	6,63	8,00	4,21	4,21	8,64	9,26
16-17	0,44	5,85	2,82	2,12	1,17	0,87	7,82	12,90	4,77	6,34	12,97	4,90
17-18	0,31	0,75	1,32	1,47	1,17	13,09	6,45	4,05	3,24	4,80	7,22	10,91
18-19	0,84	9,19	1,81	2,91	1,53	7,52	7,10	7,55	4,88	3,40	10,14	10,64
Média	0,67	3,05	1,82	2,08	1,48	3,44	4,50	10,11	5,29	5,22	8,60	9,82

Adaptado de Zhong *et al.* (2003).

Pode-se observar que as médias dos erros calculados com a aplicação dos modelos foram menores com a aplicação de regressão baseada em algoritmos, atingindo o valor máximo de 3,05 para o grupo de tráfego geral. Já os resultados obtidos através da aplicação de redes neurais chegaram a ser maiores que os modelos convencionais de imputação de dados, apresentando um valor médio de 10,11, enquanto os modelos tradicionais atingiram o valor máximo de 6,70 (modelo utilizado na França, conforme tabela).

De acordo com os autores do estudo, esta baixa eficácia do modelo baseado em redes neurais ocorreu porque não há capacidade deste método em detectar a tendência dos dados. Além disto, Zhong *et al.* (2003) também afirmam que, devido a pouca quantidade de dados, o modelo apresentou desempenho inferior, pois não é capaz de adaptar-se com baixa quantidade de dados para análise.

No geral, os autores avaliam o modelo de regressão baseado em algoritmos genéticos mais eficazes na imputação dos dados faltosos da série histórica de volume de tráfego da cidade de Alberta, Canadá. Verificando estes dados, pode-se verificar que os erros obtidos pelos modelos tradicionais não são muito superiores aos obtidos pela aplicação de redes neurais, e, portanto, podem também ser utilizados, visando uma aplicação mais fácil e rápida por parte de um pesquisador.

3.5. CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentadas as características do modelo estatístico de estimação de dados ARIMA, utilizado por alguns órgãos rodoviários e pesquisadores para imputação de dados faltosos para a variável volume de tráfego. Foram apresentados também órgãos rodoviários que ainda utilizam os modelos simples para estimação de dados faltosos, bem como a formulação matemática de alguns desses modelos. O capítulo também trata de algumas medidas de desempenho quantitativas e qualitativas, encontradas na revisão bibliográfica, as quais são utilizadas para determinar a precisão de um modelo de estimação. Por fim, foram apresentados modelos utilizados em outros estudos realizados de imputação de dados de tráfego, e os

erros obtidos com estas aplicações, buscando embasar os dados que serão encontrados neste trabalho.

O capítulo 4 a seguir inicia com uma breve apresentação dos dados dos postos de contagem contínua selecionados para o estudo, discute a escolha dos modelos ARIMA e simples, e em seguida realiza as devidas comparações dos erros estimados da aplicação dos métodos de imputação.

CAPÍTULO 4

APLICAÇÃO E AVALIAÇÃO DOS MODELOS DE ESTIMAÇÃO DE DADOS FALTOSOS

Neste capítulo serão inicialmente escolhidos os postos permanentes que serão utilizados para verificar a eficácia dos modelos de estimação de dados faltosos, sendo em seguida realizada uma caracterização dos dados válidos destes postos com relação a variação do volume mensal.

Aqui também será apresentado o processo de calibração do modelo ARIMA empregado, realizando os testes estatísticos necessários, como também serão detalhados os quatro modelos simples escolhidos. Na aplicação dos modelos escolhidos, serão retiradas amostras aleatórias dos postos selecionados com vários percentuais de dados faltosos, para comparação da eficácia dos modelos estudados com relação à quantidade de falhas para imputação e complementação da base de dados.

4.1. ESCOLHA DOS POSTOS PERMANENTES E CARACTERIZAÇÃO DO VOLUME DE TRÁFEGO

Como detalhado no Capítulo 2, o Programa de Monitoramento Contínuo do Volume de Tráfego das rodovias estaduais do Ceará possui 12 postos permanentes e apresentou, no período de funcionamento (de 1996 a 2005), altos percentuais de dados faltosos. A Figura 4.1 apresenta os percentuais de dados válidos em ordem decrescente, além dos dados faltosos e inválidos, para todos os 12 postos durante funcionamento dos equipamentos e da coleta de dados, com a formação da série histórica.

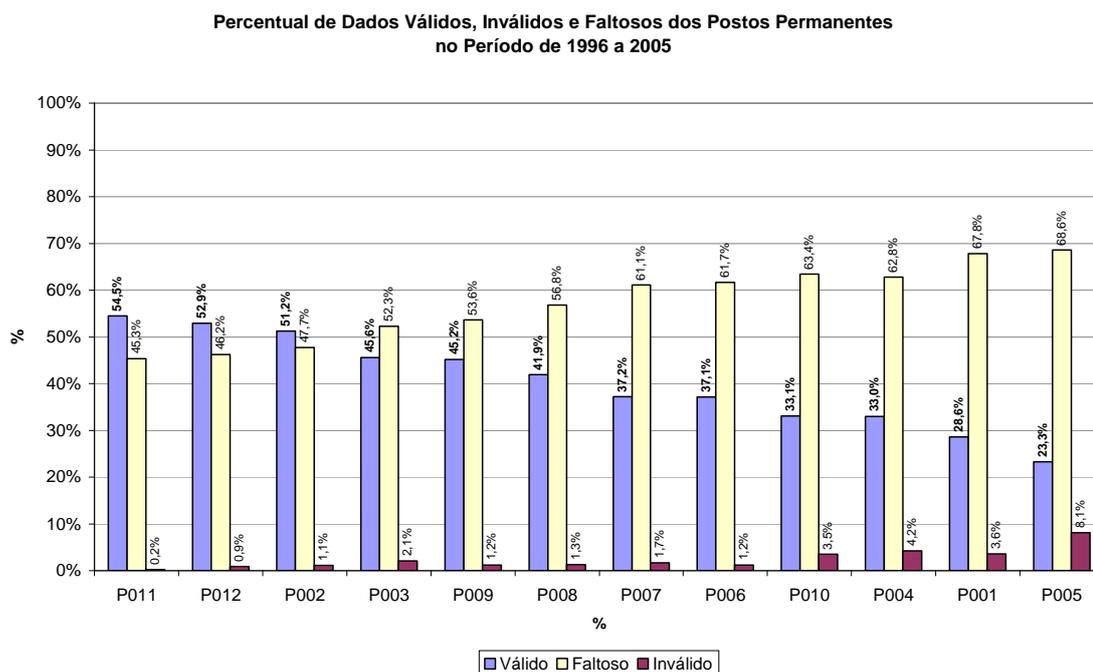


Figura 4.1: Percentual de Dados Válidos, Inválidos e Faltosos para todos os Postos de Contagem

Para a aplicação dos modelos propostos neste estudo, optou-se por utilizar postos que possuem a maior quantidade de dados válidos para o volume diário. Escolheram-se então os postos P011, P012 e P002, que conforme a Figura 4.1 apresentaram percentuais de 54,5%, 52,9% e 51,2% de dados válidos, respectivamente.

Os postos P011 e P002 estão localizados na região Norte do Estado, nas rodovias CE-354 e CE-187 respectivamente. A CE-354 é um dos principais acessos ao Litoral Oeste sendo uma importante via para o desenvolvimento do turismo dessa região. A CE-187 é uma via de integração estadual, tendo seu traçado posicionado de forma paralela ao Estado do Piauí e interliga Viçosa do Ceará (na Serra da Ibiapaba) à cidade de Crateús, seguindo até o extremo sul do Estado.

Já o posto P012 localiza-se na região Sul do Estado na rodovia CE-292. Junto com a CE-293 essa via corta transversalmente a Região Sul do Estado, fazendo a interligação da BR-116 com a BR-230, constituindo importante ligação interestadual.

Visto que os postos até agora selecionados concentram-se somente nas regiões Norte e Sul do Estado, acrescentou-se o equipamento P003, que possui um percentual

de dados válidos de somente 45,6%, mas que localiza-se na região do Sertão Central do Estado, mas especificamente na rodovia CE-060. A CE-060 é mais extensa rodovia estadual cearense com 583 km, e também denominada de “Estrada do Algodão”, cortando o Estado do Ceará de norte a sul e responsável pelo escoamento da produção agrícola dos inúmeros municípios que atravessa. A Figura 4.2 mostra a distribuição espacial dos 4 postos selecionados para o estudo.

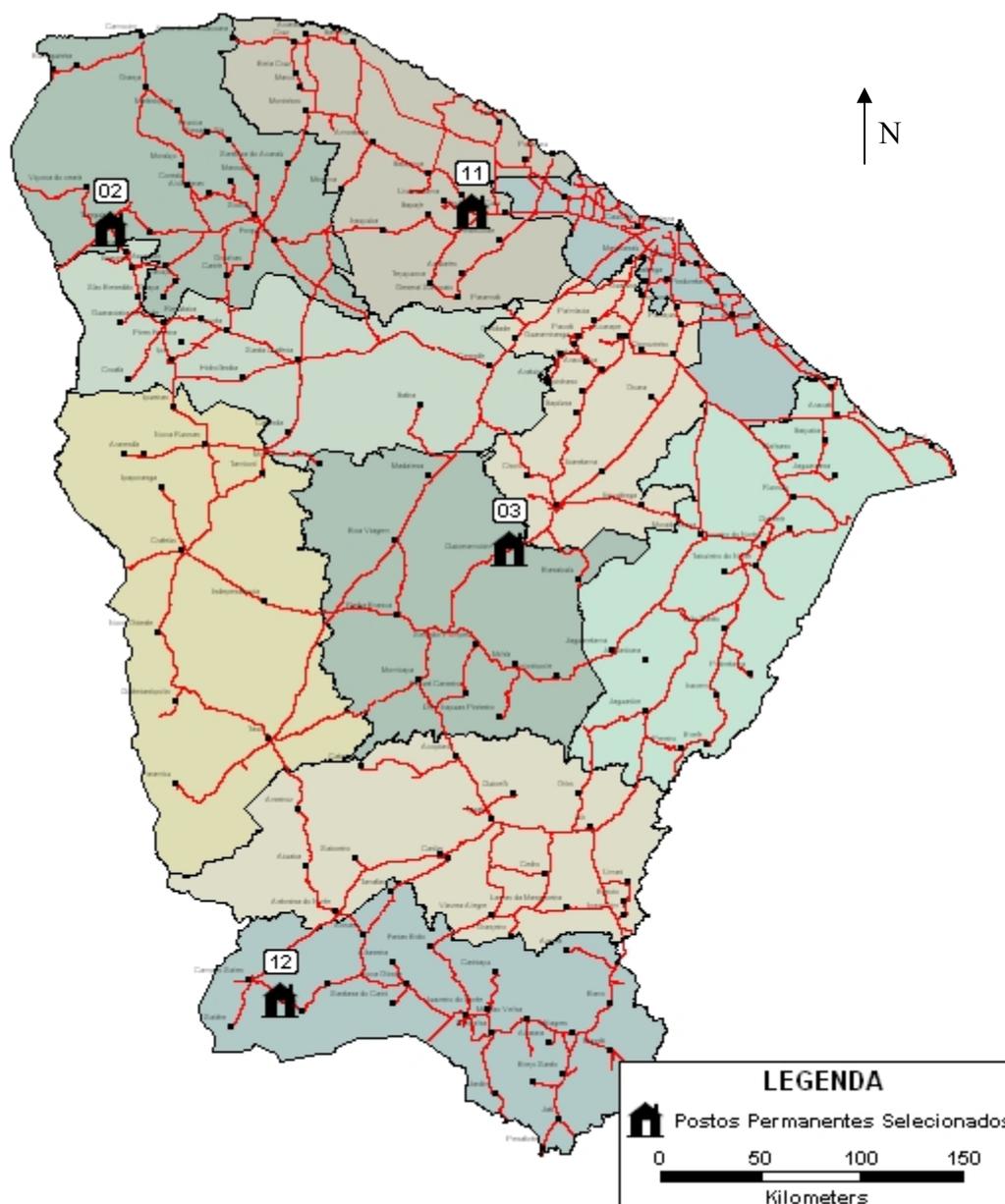


Figura 4.2: Distribuição Espacial dos Postos Permanentes Selecionados

Após a seleção dos postos, foi realizado um estudo comparativo da variação do volume de tráfego entre os equipamentos selecionados, usando-se para isso uma medida relativa, pois os postos possuem magnitudes de volume diferentes. A variável utilizada foi o Fator Mensal, que é obtida pela Equação 4.1.

$$FM = \frac{VMD_m}{VMD_a} \quad (4.1)$$

onde,

FM é o Fator Mensal;

VMD_m corresponde ao Volume Médio Diário do Mês;

VMD_a corresponde ao volume médio diário de todos os dias disponíveis do ano.

A Figura 4.3 ilustra a variação do volume de tráfego dos postos selecionados ao longo dos meses para toda a série histórica de dados. Observa-se claramente nessa figura que posto P002 não apresenta grandes variações do volume de tráfego ao longo dos meses do ano. Isso pode ser justificado pelas intensas atividades sócio-econômicas da região da Ibiapaba, no qual a via está inserida.

Nos postos P003 e P012 observam-se comportamentos semelhantes na variação mensal do volume de tráfego, influência das férias escolares, possuindo um maior pico nos meses de Julho e um aumento significativo nos meses Dezembro. Já o posto P011 apresentou as mais severas variações ao longo dos meses do ano, apresentando um pico maior nos meses de Julho e outro nos períodos de Abril e Maio, sendo esse comportamento justificado pelas viagens de características turísticas.

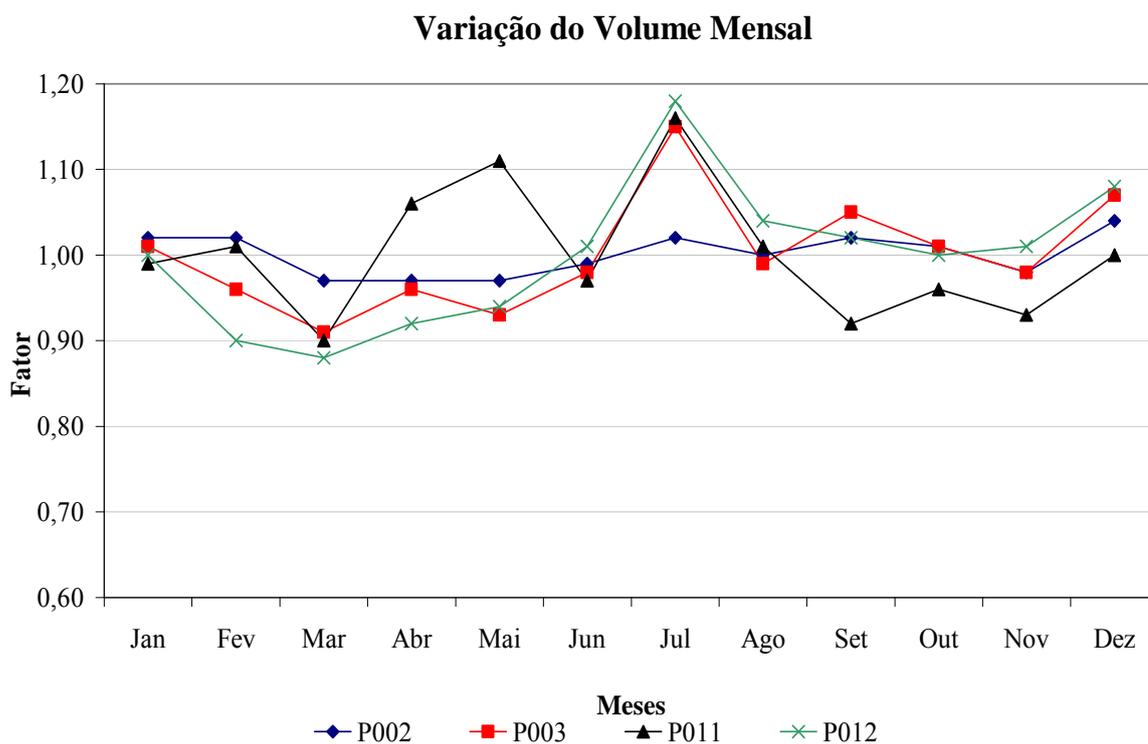


Figura 4.3: Variação do Volume de Tráfego Entre os Meses do Ano

As observações realizadas através da análise da Figura 4.3 podem ser confirmadas por meio da observação dos valores da Tabela 4.1 abaixo, que apresenta a magnitude do VMD, o desvio padrão e o coeficiente de variação dos postos permanentes selecionados neste estudo. Observe nessa tabela que os valores do coeficiente de variação apresentam são maiores para os postos P011 (30,9%), seguido, em ordem decrescente, dos postos P012 (22,1%) e P003 (18,9%), e tendo o posto P002 apresentado o menor valor (11,3%), indicando valores de volume diários bastante semelhantes ao longo do ano.

Tabela 4.1: Volume Médio Diário, Desvio Padrão e Coeficiente de Variação dos Volumes Diários para os Postos Permanentes Selecionados

POSTO	VMD	DESVIO PADRÃO	COEFICIENTE DE VARIAÇÃO
P002	1981	224	11,3%
P003	1252	236	18,9%
P011	1734	536	30,9%
P012	381	84	22,1%

4.2. IDENTIFICAÇÃO DOS MODELOS DE ESTIMAÇÃO ARIMA

A etapa de identificação do modelo ARIMA sugere inicialmente a projeção dos dados da série histórica em gráficos, para verificar a existência de algum padrão, bem como a aplicação de testes de normalidade e estacionariedade. Por isso, nessa etapa foram utilizados todos os dados válidos da série histórica dos quatro postos permanentes selecionados.

A Figura 4.4 ilustra a evolução do volume diário de veículos ao longo do período compreendido entre 1996 e 2005 no posto P002. Pode-se verificar que não há um comportamento com tendência explosiva nestes dados. O gráfico também apresenta algumas variações, e algumas lacunas sem dados que representam os períodos com dados faltosos.

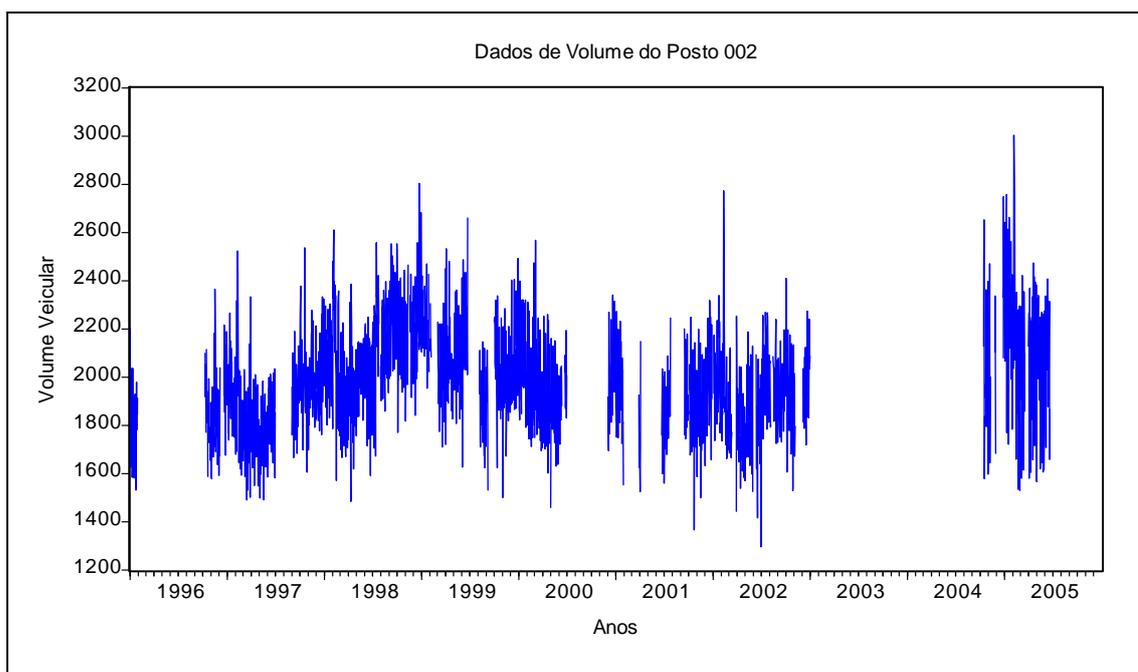


Figura 4.4: Evolução do Volume Diário de Veículos no Posto P002

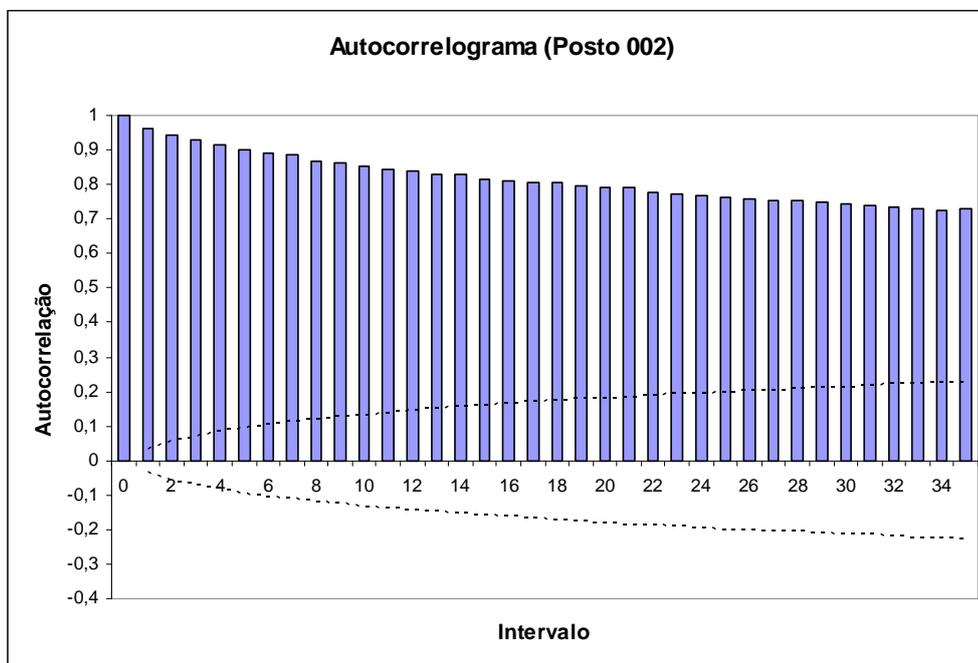
O passo seguinte foi a realização dos testes de normalidade e estacionariedade. O teste de normalidade Jarque-Bera verifica a hipótese nula H_0 : a série segue uma distribuição normal e a hipótese alternativa H_1 : a série não segue uma distribuição normal, considerando um nível de significância de 5%. Na Tabela 4.2 observando o valor-p pode-se afirmar que H_0 foi rejeitada.

Tabela 4.2: Testes de Normalidade e Estacionariedade para o Posto P002

Estatística	GL	Valor	p-valor
Jarque-Bera	2	512,651	< 0,0001
Box-Pierce	6	17665,713	< 0,0001
Ljung-Box	6	17693,393	< 0,0001
Box-Pierce	12	32951,786	< 0,0001
Ljung-Box	12	33030,201	< 0,0001

Já os testes de estacionariedade (Box-Pierce e Ljung-Box), avaliam a hipótese nula H_0 : a série possui raiz unitária (portanto é não-estacionária) e a hipótese alternativa H_1 : a série é estacionária, considerando um nível de significância de 5%. Na Tabela 4.2 observando o valor-p nas situações de 6 e 12 graus de liberdade pode-se afirmar que H_0 foi rejeitada, ou seja, a série é estacionária.

Após a realização dos testes não foi verificada a necessidade de transformação dos dados, pois a série já é estacionária, sendo o passo seguinte a identificação dos modelos potenciais através dos autocorrelograma e autocorrelograma parcial para o posto P002, como mostram as Figuras 4.5 e 4.6 respectivamente.

**Figura 4.5:** Autocorrelograma do Posto 002

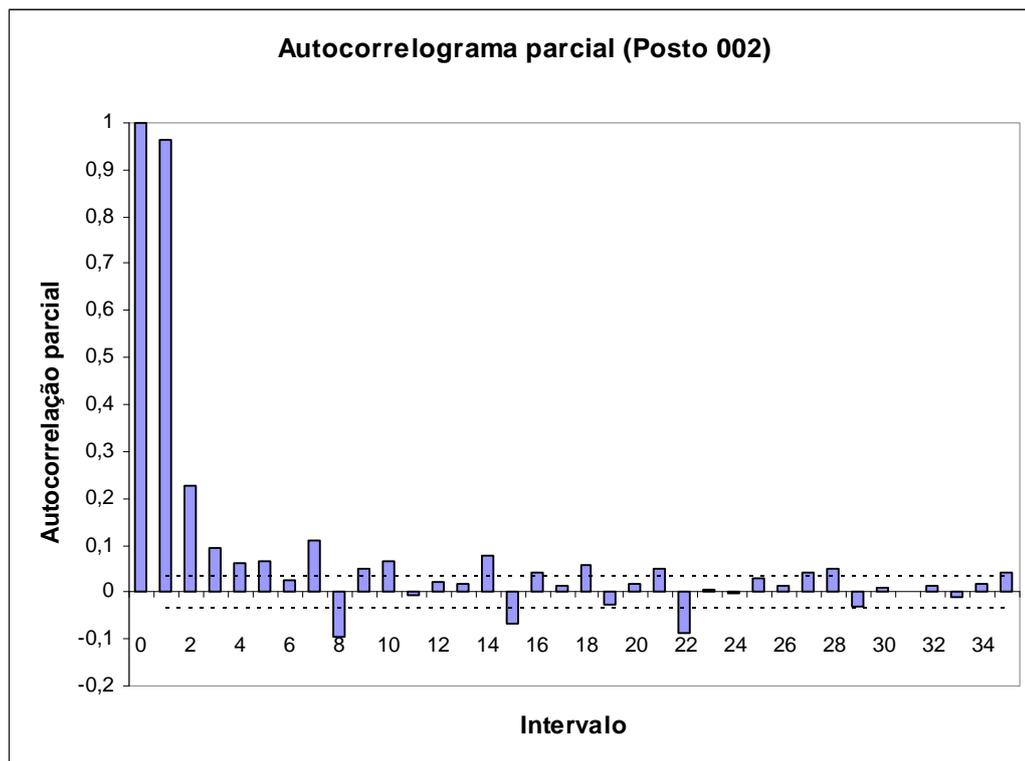


Figura 4.6: Autocorrelograma Parcial do Posto 002

Através da avaliação da Figura 4.5, pode ser verificado que não há uma tendência evidente, mas o gráfico apresenta um comportamento geométrico, declinando suavemente, o que mostra um processo do tipo AR. Já no autocorrelograma parcial ilustrado na Figura 4.6 pode-se verificar a existência de um processo MA entre os intervalos 1 e 2.

Em seguida, usando uma amostra sem dados faltosos e de tamanho 181, foram testados os 22 possíveis modelos (p,d,q) e $(P,D,Q)^S$, para o posto P002, em busca do melhor a ser utilizado na etapa de estimação. Para o caso dos modelos com componente sazonal, pode ser verificado no gráfico de autocorrelação parcial, na Figura 4.6, a existência de picos significantes a cada 7 intervalos regulares confirmando a presença de uma componente sazonal. Empiricamente essa característica sazonal é esperada, visto que se trata da variável volume de tráfego que apresenta variações semanais semelhantes.

Os modelos testados com coeficientes significativos podem ser visualizados na Tabela 4.3, bem como os parâmetros de avaliação, obtidos a partir da utilização do *software* estatístico XLStat (ADDINSOFT, 2008).

Tabela 4.3: Parâmetros Para Seleção do Modelo ARIMA no Posto 002

(p,d,q) $(P,D,Q)^S$	MAPE	AIC	BIC
(1,0,0)	6,956	2389,184	2395,581
(1,1,0)	7,497	2396,641	2403,027
(1,2,0)	10,526	2510,410	2516,785
(1,2,1)	8,026	2410,814	2420,376
(2,0,0)	6,755	2378,266	2387,862
(1,0,1) ⁷	5,842	2393,686	2441,663
(1,0,0) ⁷	6,342	2401,799	2427,387
(1,1,0) ⁷	6,411	2325,474	2331,792
(1,2,0) ⁷	9,539	2366,460	2372,696
(1,2,1) ⁷	6,385	2267,763	2277,117
(2,0,0) ⁷	6,296	2407,718	2455,696

Os parâmetros selecionados para a escolha do modelo foram o MAPE, o Critério de Informação de Akaike (AIC), e o Critério de Informação Bayesiano de Schwarz (BIC), sendo selecionados os modelos que obtiveram os menores valores nestes parâmetros. Observando a Tabela 4.3 pode-se notar que o modelo $(1,0,1)^7$ apresentou o menor valor de MAPE, e o modelo $(1,2,1)^7$ apresentou os menores valores de AIC e BIC, sendo portanto, ambos selecionados para utilização na etapa de estimação.

Para obter os modelos de estimação ARIMA dos postos P003, P011 e P012 todas essas análises foram replicadas, apresentando estes postos análises muito similares ao posto P002, sendo testados os mesmos 22 modelos para a etapa de calibração. No Anexo I (Figuras I.1 a I.3) apresenta os gráficos da evolução do volume diário de veículos ao longo da série histórica para esses equipamentos. No Anexo II (Tabela II.1 a II.3) são apresentados os demais testes de normalidade e estacionariedade, e no Anexo III (Figuras III.1 a III.6) os autocorrelogramas e autocorrelogramas parciais dos postos P003, P011 e P012.

No processo de calibração dos modelos dos postos P003, P011 e P012 utilizou-se também de amostras sem a existência de dados faltosos, com tamanhos de acordo com o ilustrado na Tabela 4.4. Os resultados dos parâmetros MAPE, AIC e BIC dos modelos com coeficientes significativos são visualizados nas Tabelas 4.5 a 4.7.

Serão considerados somente os modelos que apresentaram coeficientes significativos.

Tabela 4.4: Tamanho das Amostras de Calibração dos Modelos ARIMA Para os Postos P003, P011 e P012

	Posto 003	Posto 011	Posto 012
Tamanho das Amostras de Calibração	152	168	118

Para o posto P003, na etapa de calibração foi selecionado somente o modelo $(1,1,1)^7$, visto que esse apresentou os menores valores para os três critérios analisados (valores sublinhados), conforme mostra a Tabela 4.5.

Tabela 4.5: Parâmetros Para Seleção do Modelo ARIMA no Posto 003

$(p,d,q) (P,D,Q)^S$	<i>MAPE</i>	<i>AIC</i>	<i>BIC</i>
(1,0,0)	8,740	1931,015	1937,063
(1,1,0)	11,176	1984,715	1990,749
(1,2,0)	15,045	2074,945	2080,966
(1,2,1)	10,580	1991,786	2000,818
(2,0,0)	8,729	1927,502	1936,574
$(1,1,1)^7$	5,580	1767,976	1776,907
$(1,0,0)^7$	6,581	1876,650	1900,841
$(1,1,0)^7$	6,725	1800,281	1806,235
$(1,2,0)^7$	9,293	1809,916	1815,771

A Tabela 4.6 mostra os resultados da calibração para o posto P011, sendo escolhido para a fase de validação os modelos $(1,0,1)^7$, por apresentar o menor valor de MAPE, e o modelo $(1,2,1)^7$, que apresentou os menores valores de AIC e BIC.

Tabela 4.6: Parâmetros Para Seleção do Modelo ARIMA no Posto 011

$(p,d,q) (P,D,Q)^S$	<i>MAPE</i>	<i>AIC</i>	<i>BIC</i>
(1,0,0)	11,407	2309,066	2315,314
(1,1,0)	13,880	2340,956	2347,192
(1,2,0)	19,886	2449,760	2455,984
(1,2,1)	13,295	2347,695	2357,031
(2,0,0)	11,394	2309,977	2319,349
$(1,0,1)^7$	9,329	2302,449	2349,308
$(1,0,0)^7$	10,241	2308,919	2333,911
$(1,1,0)^7$	10,505	2227,891	2234,054
$(1,2,0)^7$	14,940	2254,342	2260,416
$(1,2,1)^7$	10,048	2169,627	2178,738
$(2,0,0)^7$	9,966	2316,915	2363,775

Por fim, os resultados da calibração para o posto P012 são mostrados na Tabela 4.7, podendo ser observado que o modelo $(1,0,1)^7$ apresentou o menor valor de MAPE, e o modelo $(1,2,1)^7$ apresentou os menores valores de AIC e BIC, sendo também os dois modelos escolhidos para aplicação na fase de validação.

Tabela 4.7: Parâmetros Para Seleção do Modelo ARIMA no Posto 012

$(p,d,q) (P,D,Q)^S$	MAPE	AIC	BIC
$(1,1,0)$	17,762	1351,347	1356,871
$(1,2,0)$	25,244	1425,555	1431,062
$(1,2,1)$	20,629	1362,290	1370,551
$(1,0,1)^7$	9,325	1268,083	1309,643
$(1,0,0)^7$	11,444	1278,807	1399,972
$(1,1,0)^7$	10,152	1184,831	1190,250
$(1,2,0)^7$	13,938	1186,527	1191,815
$(1,2,1)^7$	10,990	1150,442	1158,376
$(2,0,0)^7$	10,738	1278,911	1320,471

Na Tabela 4.8 apresenta-se um resumo dos modelos selecionados na fase de calibração para cada posto, utilizando os 2 critérios discutidos. Observa-se que para todas as amostras estudadas, os melhores modelos encontrados foram os que apresentaram componente sazonal. Também se destaca o fato que nos postos P002, P011 e P012, os modelos $(1,0,1)^7$ $(1,2,1)^7$ foram selecionados como melhores e no posto P003 o modelo $(1,1,1)^7$ foi considerado o melhor nos três critérios.

Tabela 4.8: Modelos ARIMA Selecionados Para Cada Posto de Acordo com os Critérios AIC, BIC e MAPE

Posto	MAPE	AIC e BIC
P002	$(1,0,1)^7$	$(1,2,1)^7$
P003	$(1,1,1)^7$	$(1,1,1)^7$
P011	$(1,0,1)^7$	$(1,2,1)^7$
P012	$(1,0,1)^7$	$(1,2,1)^7$

É importante frisar que, para todos os quatro postos analisados, os modelos ARIMA selecionados apresentam componente sazonal, ou seja, os valores de AIC, BIC e MAPE menores corresponderam aos modelos ARIMA $(p,d,q)^S$.

4.3. ESCOLHA DOS MODELOS SIMPLES PARA ESTIMAÇÃO DE DADOS FALTOSOS

Para verificar a eficácia na estimação de dados faltosos pelos modelos ARIMA, esses serão comparados com modelos simples de estimação. Os quatro modelos simples escolhidos foram obtidos a partir de revisão bibliográfica, podendo ter ocorrido pequenas adaptações de alguns, de maneira que fosse permitida a obtenção dos valores faltosos, utilizando-se somente de planilha eletrônica, sem a necessidade de cálculos complexos. Os modelos escolhidos estão descritos a seguir.

a) *Modelo Simples 1 (MS 1): Média Histórica*

O valor perdido é obtido a partir da média dos dados históricos de todos os anos anteriores, considerando o mesmo dia da semana, na mesma semana do mês e no mesmo mês da falha existente. Considere o exemplo de estimar um valor para o volume diário da data de 11/07/1999. Essa data é um Domingo e pertence a terceira semana do mês de Julho. Assim, identifica-se os Domingos pertencentes a terceira semana do mês de Julho para os anos anteriores a falha (1998, 1997, 1996) e calcula-se a média desses dados como estimativa para o valor faltoso do dia 11/07/1999.

Uma limitação na utilização desse primeiro modelo é que para o ano de 1996, não foi possível realizar estimação dos dados faltosos, visto que não havia contagens contínuas anteriores a esse ano. A estimação de dados faltosos para o ano de 1997 também foi limitada visto que, o ano base para esses cálculos é o ano de 1996 que possui somente 32% de dados válidos. Essa razão prejudica a maioria dos modelos de estimação simples, visto que se utiliza nos cálculos os dados dos anos anteriores, e caso não existam dados para esses dias a estimação do dado faltoso será nula.

b) *Modelo Simples 2 (MS 2): Média Antes e Depois*

Esse modelo possui o mesmo princípio de cálculo do anterior, porém tenta melhorar os resultados de estimação a partir da média histórica entre os anos antes e

depois da falha. Neste caso, se fosse necessário estimar o valor do volume diário para a data 11/07/1999, realizar-se-ia a média dos Domingos pertencentes a terceira semana do mês de Julho para os anos anteriores (1998, 1997, 1996) e para os anos posteriores (2000, 2001, 2002, 2003, 2004 e 2005).

Esse modelo reduz as dificuldades encontradas pelo modelo MS 1, visto que a média é obtida por dados históricos dos anos antes e após a falha, sendo mais provável a existência de pelo menos um valor em toda a série histórica que possa ser utilizado.

c) Modelo Simples 3 (MS 3): Interpolação/Mês

Nesse terceiro modelo, o valor perdido é estimado a partir da média dos dados do mês anterior e do mês posterior a falha, considerando o mesmo dia da semana e na mesma semana do mês. Assim, caso fosse necessário estimar um valor do volume diário para a data 11/07/1999, aplicar-se-ia a Equação 4.2:

$$MS3 = \frac{FM_{i-1} * valor_{i-1} + FM_{i+1} * valor_{i+1}}{2} \quad (4.2)$$

onde,

FM_{i-1} é o Fator Mensal do mês anterior à falha;

$valor_{i-1}$ é o valor do mês anterior à falha;

FM_{i+1} é o Fator Mensal do mês posterior à falha;

$valor_{i+1}$ é o valor do mês posterior à falha.

Este modelo, porém, apresenta um ponto negativo, que se assemelha ao apresentado pelo MS1, pois, se o dado do mês anterior ou do mês posterior foi perdido, a estimação não ocorre.

d) Modelo Simples 4 (MS 4): Interpolação/Ano

Nesse modelo, o valor perdido é estimado a partir da média dos dados do ano anterior e do ano posterior a falha, considerando o mesmo dia da semana, a mesma

semana do mês e o mesmo mês da falha existente. Assim, caso fosse necessário estimar um valor do volume diário para a data 11/07/1999, aplicaria-se a Equação 4.3.

$$MS4 = \frac{\left(\frac{VMDA_{antes}}{VMDA_{falha}}\right) * valor_{antes} + \left(\frac{VMDA_{depois}}{VMDA_{falha}}\right) * valor_{depois}}{2} \quad (4.3)$$

onde,

$VMDA_{antes}$ é o Volume Médio Diário Anual do ano anterior à falha;

$VMDA_{falha}$ é o Volume Médio Diário Anual do ano da falha;

$VMDA_{depois}$ é o Volume Médio Diário Anual do ano posterior à falha;

$valor_{antes}$ é o valor do volume do ano anterior à falha;

$valor_{depois}$ é o valor do volume do ano posterior à falha.

4.4. APLICAÇÃO E ANÁLISE DOS MODELOS DE ESTIMAÇÃO

Após a escolha dos modelos foram selecionadas amostras aleatórias em cada um dos 4 postos permanentes (P002, P003, P011 e P012), sendo utilizadas tanto para a aplicação dos modelos ARIMA quanto para os modelos simples. O objetivo foi a comparação da eficácia desses na estimação de dados faltosos. Nessa etapa as amostras extraídas foram diferentes das amostras retiradas para calibração dos modelos ARIMA. As amostras foram retiradas da série história de intervalos que não apresentavam falhas (por isso foram selecionados os equipamentos com maior percentual de dados válidos), que foram apagadas de acordo com a necessidade de dados faltosos. Ou seja, para avaliar uma amostra com 5% de dados faltosos, utilizaram-se dados sem falhas, e foram retirados aleatoriamente 5% dos dados.

Para analisar uma amostra com 10% de dados faltosos foi selecionada outra amostra sem falhas, e foram apagados 10% de dados. Ou seja, foram selecionados seis conjuntos de amostras sem falhas, e de cada amostra foi retirado aleatoriamente um percentual de dados para a aplicação dos modelos. Assim, a Tabela 4.9 apresenta as amostras selecionadas, com seus respectivos tamanhos e relação com o percentual dados que foram apagados, de acordo com a demanda para avaliação.

Tabela 4.9: Tamanho das Amostras de Cada Posto Para Aplicação dos Modelos

<i>Percentual de Dados Faltosos</i>	Tamanho das Amostras por Posto			
	Posto 002	Posto 003	Posto 011	Posto 012
5%	104	105	87	82
10%	88	89	90	82
20%	100	132	110	82
30%	104	153	132	90
40%	85	193	155	90
50%	184	82	170	118

Realizada a aplicação de todos os modelos, foi utilizado o parâmetro MAPE como critério para verificar o ajuste dos dados estimados com relação aos dados da amostra. A Tabela 4.10 ilustra os resultados desse parâmetro para todos os modelos e amostras com diferentes percentuais de dados faltosos, retirados do posto P002.

Tabela 4.10: Erros tipo MAPE obtidos com a aplicação dos modelos em relação ao percentual de dados perdidos do Posto 002

	$(1,0,1)^7$	$(1,2,1)^7$	<i>MS 1</i>	<i>MS 2</i>	<i>MS 3</i>	<i>MS 4</i>
5%	0,563	0,553	2,996	0,317	1,294	3,020
10%	0,592	0,912	10,227	0,776	3,995	10,227
20%	1,263	1,596	4,676	0,937	5,550	4,965
30%	1,502	2,020	15,777	2,542	8,019	20,784
40%	2,804	3,068	4,048	3,270	13,525	31,103
50%	4,174	4,926	5,455	3,191	9,689	38,076
Média	1,816	2,179	7,197	1,839	7,012	18,029

Pela análise dos valores da Tabela 4.10, três dos modelos simples obtiveram variações maiores do MAPE, tendo o MS 1 valores mais próximos, variando entre aproximadamente 3% e 15,7%, o MS 3 entre 1,3% e 13,5% e o MS 4 valores entre 3,0% e 38,1%. Nos demais modelos as variações foram menores, tendo o MS 2 apresentado erros de no máximo 3,27% e os modelos ARIMA erros inferiores a 5%. Considerando os menores valores obtidos para cada amostra, o modelo $(1,0,1)^7$ apresentou em 3 casos superioridade com relação aos demais, bem como o modelo MS 2. A escolha do modelo mais eficaz levou em consideração, portanto, a menor variação dos erros, bem como os valores da média das amostras para cada modelo aplicado, sendo, portanto o modelo ARIMA $(1,0,1)^7$ identificado como mais eficaz para estimação dos dados faltosos do posto P002, seguido do MS 2, que apresentou um erro médio um pouco maior (com diferença de 0,023).

A Tabela 4.11 apresenta os erros MAPE obtidos com a aplicação dos modelos propostos para os dados do Posto 003.

Tabela 4.11: Erros tipo MAPE obtidos com a aplicação de cada modelo com relação ao percentual de dados perdidos do Posto 003

	$(1,1,1)^7$	MS 1	MS 2	MS 3	MS 4
5%	0,299	2,268	0,465	0,710	4,762
10%	1,458	1,606	1,052	7,677	8,571
20%	1,873	24,762	2,221	12,843	24,762
30%	2,205	11,104	2,485	12,515	25,675
40%	5,876	29,687	4,837	28,956	50,877
50%	4,289	3,426	2,596	20,713	39,048
Média	2,667	12,142	2,276	13,902	25,616

Na análise dos valores da Tabela 4.11, três dos modelos simples obtiveram variações maiores do MAPE, tendo o MS 1 valores aproximados entre 1,6% e 29,7%, o MS 3 entre 0,7% e 28,9% e o MS 4 valores 4,7% e 50,9%. Nos demais modelos as variações foram menores, tendo o MS 2 apresentado erros abaixo de 5% e o modelo ARIMA erros inferiores a 6%. No caso dos dados obtidos pelo modelo ARIMA $(1,1,1)^7$ e pelo MS 2, em três casos apresentaram superioridade com relação aos demais, por possuir os menores erros nestes casos. Para a escolha do modelo mais eficaz levou-se em consideração, portanto, a análise dos menores erros, bem como os valores da média das amostras para cada modelo aplicado, sendo, portanto o MS 2 identificado como mais eficaz para estimação dos dados faltosos do posto P003, por ter apresentado também o menor valor médio de MAPE (2,27%), seguido do ARIMA $(1,1,1)^7$ com o valor médio de 2,67%.

A Tabela 4.12 a seguir apresenta os erros MAPE obtidos com a aplicação dos modelos propostos para análise dos resultados obtidos para o posto P011.

Tabela 4.12: Erros tipo MAPE obtidos com a aplicação de cada modelo com relação ao percentual de dados perdidos do Posto 011

	$(1,0,1)^7$	$(1,2,1)^7$	MS 1	MS 2	MS 3	MS 4
5%	0,558	0,599	0,666	0,994	0,697	4,598
10%	1,215	1,096	4,119	3,223	6,063	10,000
20%	2,854	3,409	3,950	6,163	10,192	6,709
30%	3,794	4,225	25,457	5,300	8,426	25,258
40%	2,974	3,224	9,980	5,000	17,519	11,054
50%	8,319	10,618	8,114	7,644	17,852	21,230
Média	3,286	3,862	8,714	4,721	10,125	13,141

Através da verificação dos valores da Tabela 4.12, pode-se notar que três dos modelos simples obtiveram variações maiores do MAPE, tendo o MS 1 valores aproximados entre 0,67% e 25,5%, o MS 3 entre 0,69% e 17,8% e o MS 4 valores 4,6% e 25,26%. Nos demais modelos as variações foram menores, tendo o MS 2 apresentado erros inferiores a 8%, e os modelos ARIMA erros inferiores a 11%. Analisando os valores obtidos pelo modelo ARIMA $(1,0,1)^7$, em cinco casos apresentou superioridade com relação aos demais, por possuírem menores erros.

Portanto, para a escolha do modelo mais eficaz, levou-se em consideração a análise dos menores erros, bem como os valores da média das amostras para cada modelo aplicado, sendo o $(1,0,1)^7$ identificado como mais eficaz para estimação dos dados faltosos do posto P011, por ter apresentado também o menor valor médio de MAPE (3,286%), seguido pelos modelos ARIMA $(1,2,1)^7$, com uma valor médio de MAPE 3,862% e o MS 2, com um MAPE médio de 4,721%.

A Tabela 4.13 apresenta os erros MAPE obtidos com a aplicação dos modelos nos dados do posto P012, para análise dos resultados obtidos.

Tabela 4.13: Erros tipo MAPE obtidos com a aplicação de cada modelo com relação ao percentual de dados perdidos do Posto 012

	$(1,0,1)^7$	$(1,2,1)^7$	MS 1	MS 2	MS 3	MS 4
5%	0,302	0,561	0,399	0,232	2,699	3,791
10%	1,731	2,454	2,051	1,275	2,376	7,637
20%	3,321	3,756	5,104	1,975	5,161	8,428
30%	4,414	8,566	15,734	2,985	14,210	15,645
40%	4,443	5,557	40,000	7,208	21,020	40,000
50%	8,277	9,230	29,485	4,663	25,116	36,703
Média	3,748	5,021	15,462	3,056	11,764	18,701

Através da verificação dos valores da Tabela 4.13, pode-se notar que três dos modelos simples obtiveram variações maiores do MAPE, tendo o MS 1 valores aproximados entre 0,4% e 40%, o MS 3 entre 2,4% e 25,12% e o MS 4 valores 3,79% e 40%. Nos demais modelos as variações foram menores, tendo o MS 2 apresentado erros inferiores a 8%, e os modelos ARIMA erros inferiores a 10%. Analisando os dados obtidos pelo MS 2, em cinco casos apresentou superioridade com relação aos demais por possuírem os menores erros.

Já o modelo ARIMA (1,0,1)⁷ apresentou somente um caso com o menor erro do tipo MAPE com relação aos demais modelos, quando ocorre a falta de 40% dos dados. Para a escolha do modelo mais eficaz levou-se em consideração, a análise dos menores erros, bem como os valores da média das amostras para cada modelo aplicado, sendo o MS 2 identificado como o que possuiu mais casos de menores erros e o que apresenta a menor média dos MAPE's (3,056%) sendo considerado, portanto, o modelo mais eficaz para estimação dos dados faltosos do posto P012, seguido do modelo ARIMA (1,0,1)⁷ com um valor médio de 3,748% de MAPE.

A Tabela 4.14 apresenta um resumo dos modelos ARIMA e Simples mais eficazes para cada um dos postos permanentes analisados neste estudo.

Tabela 4.14: Resumo dos Erros dos Modelos Avaliados como Eficazes

<i>Posto</i>	<i>ARIMA</i>	<i>Modelo Simples</i>
P002	(1,0,1) ⁷ 1,816%	MS 2 1,837%
P003	(1,1,1) ⁷ 2,667%	MS 2 2,276%
P011	(1,0,1) ⁷ 3,286%	MS 2 4,721%
P012	(1,0,1) ⁷ 3,748%	MS 2 3,056%

Pode-se notar pela Tabela 4.14, que o modelo MS 2 apresentou melhores resultados para os postos P003 e P012, e para os postos P002 e P011 o modelo ARIMA (1,0,1)⁷ foi considerado o melhor modelo de estimação. Para esse último caso, observa-se que os erros médios calculados entre as amostras foram muito próximos, tendo no posto P002 o ARIMA obtido o valor de 1,816% e o MS 2 de 1,837% e no P011 os percentuais foram 3,748% do ARIMA contra 4,721% do MS 2.

Portanto, a adoção do modelo simples MS 2 para estimação de dados faltosos de volume de tráfego nesses 4 postos permanentes é tão viável quanto a utilização de modelos como o ARIMA, tendo as vantagens de sua aplicação ser menos complexa e mais rápida. Conforme visto no capítulo anterior, outros autores também consideraram outros modelos de estimação que, mesmo não obtendo menores erros, estavam próximos dos menores.

4.5. CONSIDERAÇÕES FINAIS

Neste capítulo foi realizada a escolha dos postos de contagem contínua do PMT do Estado do Ceará e em seguida a caracterização destes quanto a variação mensal do volume de tráfego. Após a escolha dos postos para aplicação dos modelos de imputação, realizou-se a etapa de calibração dos modelos ARIMA, sendo identificado 2 modelos ARIMA para aplicação nos dados dos postos P002, P011 e P012 e um modelo para aplicação no posto P003. Utilizou-se também de quatro modelos simples permitindo comparar a eficácia dos modelos ARIMA na estimação de dados faltosos em relação a estes modelos menos complexos. Para a aplicação de todos os modelos foram selecionadas amostras das séries de dados dos postos selecionados, levando em consideração diferentes percentuais de dados faltosos, de 5 a 50%.

A aplicação e análise dos resultados demonstraram que o modelo simples MS 2 apresentou em 2 postos resultados melhores que os modelos ARIMA (P003 e P012) e em outros 2 postos os resultados do modelo ARIMA, apesar de possuírem menores erros médios, foram muito próximos dos obtidos pelo MS 2 (P002 e P011), sendo portanto esse modelo simples também considerado eficaz na estimação de dados faltosos para esses 4 postos.

CAPÍTULO 5

CONCLUSÕES E RECOMENDAÇÕES

5.1. CONCLUSÕES

A principal contribuição do presente trabalho foi a investigação sobre a eficácia dos modelos simples e modelos de série temporais ARIMA, para a estimação de dados faltosos na série histórica (1996 a 2005) do volume de tráfego dos equipamentos de contagem contínua, instalados nas rodovias estaduais do Ceará. Embora em revisão bibliográfica tenham sido encontrados vários trabalhos ressaltando a precisão do modelo ARIMA na estimação de dados faltosos, o banco de dados existente para as rodovias estaduais apresenta um percentual de dados faltosos muito maior que em outros estudos, o que pode gerar problemas para obter essa melhor precisão.

Assim, vale ressaltar que esta pesquisa consistiu em um esforço de caráter confirmatório, sobre a adoção de modelos ARIMA na estimação de dados faltosos apresentar resultados melhores que modelos simples, sendo que na prática esses métodos mais simples ainda são muito utilizados pelos órgãos rodoviários, como pode ser observado em revisão bibliográfica.

Para tanto, foi estudada a estruturação de um programa de monitoramento de tráfego rodoviário, e foram apresentadas algumas formas de identificação e controle de inconsistências na base de dados, sendo permitido manter a qualidade das informações utilizadas como subsídios básicos em vários estudos de planejamento e operação rodoviária. Sabe-se que é inapropriado exigir 100% de qualidade para os dados coletados pelos detectores dos postos permanentes, mas alguns procedimentos podem influenciar muito na melhoria da qualidade dos dados, tais como: manutenção dos equipamentos de detecção, sistema de monitoramento de falhas, e principalmente disponibilidade de um sistema de comunicação para a transferência de dados por meio remoto.

Em seguida foram descritos e detalhados os modelos ARIMA e os principais modelos simples utilizados para estimação de dados faltosos por alguns órgãos rodoviários. Discutiram-se também algumas medidas de desempenho quantitativas e qualitativas, as quais podem ser utilizadas isoladamente ou em conjunto para verificar a precisão de uma modelo de estimação.

No processo de calibração do ARIMA foram utilizados 3 critérios (MAPE, AIC e BIC) permitindo selecionar dois modelos $(1,0,1)^7$ e $(1,2,1)^7$ para os postos P002, P011 e P012 e para o posto P003 os três critérios indicaram o modelo $(1,1,1)^7$. Os três modelos ARIMA escolhidos para aplicação apresentam componente sazonal, resultado intuitivamente esperado e semelhante a outros estudos.

No que se refere ao resultado final da aplicação dos modelos propostos, a análise dos resultados demonstraram que o modelo simples do tipo MS 2 apresentou em dois postos (P003 e P012) resultados melhores que os modelos ARIMA, e nos outros dois postos (P002 e P011) os resultados do modelo ARIMA, apesar de possuírem menores erros médios, foram muito próximos dos obtidos pelo MS 2. Os erros médios das amostras do MS 2 foram os menores dos quatro modelos simples propostos, tendo valor mínimo de 1,837% no posto P002 e máximo de 4,721% para o posto P011. Os erros médios do ARIMA variaram entre 1,816% no P002 e 3,748% para o P012. Portanto, pode-se concluir que o MS 2, mostrou-se tão bom quanto o ARIMA na estimação de dados faltosos aplicados nestes 4 postos de contagem.

Uma desvantagem do uso de modelos estatísticos do tipo ARIMA é que o processo de calibração e aplicação são complexos, com análises subjetivas e demoradas, além de ser de difícil implementação fora de um *software* estatístico, necessitando o órgão rodoviário de um profissional técnico altamente especializado na sua utilização.

Os quatro modelos simples escolhidos para aplicação são bastante intuitivos, sendo necessário somente um conhecimento intermediário de planilha eletrônica, podendo facilmente implementar rotinas baseadas nesses modelos para geração automática de dados faltosos dentro do banco de dados do programa de monitoramento de tráfego.

5.2. RECOMENDAÇÕES

Cabe salientar, que este trabalho não pretende esgotar as discussões sobre a temática de modelos de estimação de dados faltosos de volume de tráfego, sendo identificadas algumas pesquisas futuras que possam contribuir, considerando as seguintes recomendações:

- a) Aplicar outros modelos simples propostos pela literatura, buscando conhecer se outro método apresenta eficácia maior ou igual aos encontrados neste estudo;
- b) Aplicar os modelos de estimação aqui propostos nos demais postos de contagem contínua do programa de monitoramento do volume de tráfego do Estado do Ceará, buscando conhecer o comportamento desses modelos em postos com maiores percentuais de dados faltosos.
- c) Utilizar uma série histórica completa, sem dados faltosos, e simular a perda de dados, gerando amostras com diferentes quantidades de dados faltosos para aplicação dos modelos de estimação propostos e realizar a comparação dos valores estimados com os valores reais retirados.
- d) Aplicar outros métodos estatísticos para estimar dos dados faltosos, e comparar com os modelos ARIMA e modelos simples, buscando obter os modelos mais eficientes.

REFERÊNCIAS

- AASHTO (1992) *Guidelines for Traffic Data Programs*. American Association of State Highway and Transportation Officials. Washington, D.C.
- ADDINSOFT (2008) XLStat data analysis and statistics add-in for MS Excel. Paris, France.
- AÊREDES, A. F. e PEREIRA, M. W. G. (2008) *Comportamento do Preço do Trigo no Paraná, 1994 a 2007*. Informações Econômicas, São Paulo, volume 38, número 1, pp.21-31.
- AKAIKE, H. (1973) *Information Theory and an Extension of the Maximum Likelihood Principle*. In: B. N. PETROV and F. CSAKI, eds. Second International Symposium on Information Theory. Budapest: Akademiai Kiado, pp. 267–281.
- ALBRIGHT, D. (1991a) History of Estimating and Evaluating Annual Traffic Volume Statistics. Transportation Research Record 1305, Transportation Research Board, Washington, D.C, pp. 103-107.
- ALBRIGHT, D. (1991b) An Imperative for, and Current Progress toward, National Traffic Monitoring Standards. ITE Journal, Vol. 61, No. 6, 1991, pp. 23-26.
- ALBRIGHT, D. (1993) Standards, Innovation, and the Future of Traffic Monitoring. *Institute of Transportation Engineers Journal*, Vol. 63, Nº 1, pp. 31-36.
- ASTEF (1994) *Plano de Contagem de Tráfego - Relatório Final*. Associação Técnico-Científico Eng. Paulo de Frocin, Universidade Federal do Ceará, Fortaleza, Ceará.
- ASTEF (2001) *Reestruturação do Sistema de Contagem de Tráfego para as Rodovias Estaduais – Proposta Técnica e Comercial*. Associação Técnico-Científico Eng. Paulo de Frocin, Universidade Federal do Ceará, Fortaleza, Ceará.
- ASTM (1991) Standard Practice E1442, Highway Traffic Monitoring Standards. American Society for Testing and Materials (ASTM), Philadelphia, PA.
- BEZERRA, M. I. S. (2006) *Apostila de Análise de Séries Temporais*. Curso de Estatística. DMEC / FCT / UNESP.
- BOX, G. E. P. e JENKINS, G. M. (1976) *Time Series Analysis: forecasting and control*. San Francisco, Holden-Day.
- BOX, G. E. P.; JENKINS, G. M. e REINSEL, G. (1994) *Time Series Analysis: Forecasting and Control*. 3.^a edição Hardcover, New York.
- CHEN, Y.; XIA, B. J e LIU, R. R. (2006) Developing a Strategy for Imputing Missing Traffic Volume Data. *Journal of the TRF*, Volume 45, Número 3.

- CNT (2006) Confederação Nacional dos Transportes. *Atlas do Transporte*. 1ª Edição.
- CORDEIRO, D. M. (2002) *Séries Temporais – Análise Quantitativa, Teoria e Aplicações*. Editora Universidade de Pernambuco. Recife, PE.
- DERT (2006) *Rede Rodoviária Estadual Pavimentada a ser considerada no cálculo dos recursos da CIDE - Extensões totais em 31/10/2006*. Departamento de Edificações e Rodovias do Estado do Ceará. Disponível em: www.dert.ce.gov.br. Acesso em 10/10/2008.
- DICKEY, D.A. e FULLER W. A. (1979) Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, pp. 427–431.
- DINIZ, H.; ANDRADE, L. C. M.; CARVALHO, A. C. P. L. F. e ANDRADE, M. G. (1998) Previsão de séries temporais utilizando redes neurais artificiais e modelos de Box & Jenkins. *Anais do Simpósio Brasileiro de Redes Neurais* (pp. 173-178); Belo Horizonte, MG.
- DNER (1976) *Manual de Contagem Volumétrica de Tráfego*. Departamento Nacional de Estradas e Rodagem, Rio de Janeiro, RJ.
- DNIT (2006) *Manual de Estudos de Tráfego*. Departamento Nacional de Infra-Estrutura de Transportes, Rio de Janeiro, RJ.
- DNIT (2008) *Programa Nacional de Contagem de Tráfego*. Departamento Nacional de Infraestrutura em Transportes. Disponível no endereço eletrônico: http://www1.dnit.gov.br/rodovias/contagem/plano_contagem.htm. Acesso em 12/09/2008. Brasília, DF.
- EHLERS, R. S. (2007) *Análise de Séries Temporais*. Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná.
- ENDERS, w. (1995) *Applied Econometric Time Series*. New York: John Wiley & Sons Inc.
- ENGEVIAS (2009) *Relatório de Programa de Tráfego ao Departamento de Edificações e Rodovias do Ceará*. Engenharia e Consultoria LTDA. Fortaleza, Ceará.
- FHWA (1997) *Study Tour for European Traffic Monitoring Programs and Technologies*. Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., Agosto de 1997.
- FHWA (2001) *Traffic Monitoring Guide*. Federal Highway Administration (2001), U.S. Departamento of Transportation, Washington, D.C., Maio de 2001.

- GARCIA, C. A. F. M. (2000) Análise de Séries Temporais com Recursos a Técnicas de Bases de Dados. *Dissertação de Mestrado*. Mestrado em Inteligência Artificial e Computação. Faculdade de Engenharia da Universidade do Porto. Portugal.
- GEIPOT (1970) *Contagem Normal de Tráfego*. Empresa Brasileira de Planejamento dos Transportes, Brasília, DF.
- GOMES, M. J. T. L. (2004) Volume Horário de Projeto para as Rodovias Estaduais do Ceará: Análise e Contribuição. *Dissertação de Mestrado*. Programa de Mestrado em Engenharia de Transportes. Universidade Federal do Ceará.
- GRANGER, C. W. J. e NEWBOLD, J. P. (1976) *Forecasting Transformed Series*. *Journal of the Royal Statistical Society, Series B*, 38, pp. 189-203.
- GUIMARÃES, A. C. T. (2001) Diagnóstico Preliminar da Base de Dados de Tráfego da Malha Rodoviária Estadual da Paraíba. *Dissertação de Mestrado*. Universidade Federal da Paraíba, Paraíba.
- GUJARATI, D. N. (2000) *Econometria Básica*. Trad. Ernesto Yoshida. 3ª edição. São Paulo: Makron Books.
- MAKRIDAKIS, S. e HIBON, M. (1979) Accuracy of Forecasting: An Empirical Investigation. *Journal of the Royal Statistical Society, Series A*, 142, pp. 97-145.
- MCSHANE, W.R e ROESS R. P. (1990) *Traffic Engineering*. Prentice Hall, Englewood Cliffs, New Jersey, USA.
- MOHAMAD, D.; SINHA, K. C.; KUCZEK, T. e SCHOLER, C. F. (1998) Annual Average Daily Traffic Prediction Model for County Roads. *Transportation Research Board, The 77th Annual Meeting*, Washington, D.C., USA.
- MOREIRA, M. E. P. e LIMA L. C. (1995) *Plano de Contagem de Tráfego na Malha Rodoviária do Estado do Ceará*. IX Congresso de Pesquisa e Ensino em Transportes, São Carlos, São Paulo.
- MORETTIN, P. A. (1999) *Ondas e Ondaletas. Da Análise de Fourier à Análise de Ondaletas*. São Paulo: Editora da Universidade de São Paulo.
- MORETTIN, P. A., TOLOI, C. M. (2006) *Análise de Séries Temporais*. ABE. Projeto Fisher. Edgard Blücher. São Paulo.
- NELSON, H. L. (1976) The Use of Box-Cox Transformations in Economic Time Series Analysis: An Empirical Study. *Dissertação de Doutorado*, Universidade da Califórnia, San Diego.
- OLIVEIRA, M. V. T. (2004) A Natureza dos Padrões de Variação Espaço-Temporal do Volume Veicular em Ambiente Urbano: Estudo de Caso em Fortaleza. *Dissertação de Mestrado*. Universidade Federal do Ceará, Ceará.

- PLOSSER, C. I. (1979) *A Time Series Analysis of Seasonality in Econometric Models*. NBER Chapters em Seasonal Analysis of Economic Research, Inc.
- SCHWARZ, Gideon (1978) *Estimating the Dimension of a Model*. The Annals of Statistics, 6(2), pp. 461–464.
- SHARMA, S. C.; P. LINGRAS e M. ZHONG (2003) Effect of Missing Value Imputations on Traffic Parameters Estimations from Permanent Traffic Counts. *Transportation Research Board*, The 82th Annual Meeting, Washington, D.C.
- SHARMA, S. C. e WERNER, A. (1981) Improved Method of Grouping Provincewide Permanent Traffic Counters. *Transportation Research Record 815*, Transportation Research Board, Washington, D.C.
- SHASHA, D. (1999) *Time Series in Finance: the array database approach*. ACM SIGMOD, Abril.
- SILVA, A. F. (2005) Tendências Sobre o Número de Passageiros Transportados em Ônibus Interestaduais no Brasil. *Tese de Doutorado*. Escola de Engenharia de São Carlos, Universidade de São Paulo. São Paulo, SP.
- SMITH, B. L.; W. T. SCHERER e J. H. CONKLIN (2003) Exploring Imputation Techniques for Missing Data in Transportation Management Systems. *Transportation Research Board*, The 82th Annual Meeting, Washington, D.C., USA.
- STEVENSON, J. S. (1981) *Estatística Aplicada à Administração*. São Paulo: Harper and How do Brasil.
- THEIL, H. (1966) *Applied Economic Forecasting*. North Holland.
- TURNER, S. M. (2001) *Guidelines for Developing ITS Data Archiving Systems*. Report 2127-3. Texas Department of Transportation.
- ZHONG, M.; LINGRAS, P. e SHARMA, S. C. (2002) *Updating Missing Values of Traffic Counts: Factor Approaches, Time Series Analysis versus Genetically Designed Regression and Neural Network Models*. Department Report, Department of Mathematics and Computing Science, Saint Mary's University, Halifax.
- ZHONG, M.; SHARMA, S. C. e LINGRAS, P. (2004) “Genetically Designed Models for Accurate Imputations of Missing Traffic Counts”. *Transportation Research Records 1879*, Journal of Transportation Research Board, Washington DC, pp. 71-79.
- ZHONG, M.; SHARMA, S. C. e LINGRAS, P. (2006) Matching Patterns for Updating Missing Values of Traffic Counts. *Transportation Planning and Technology*, volume 29, número 2, pp. 141-156.

ZHONG, M; SHARMA, S. C. e LIU, Z. (2005) Assessing Robustness of Imputation Models based on Data from Different Jurisdictions. Transportation Research Board: *Journal of the Transportation Research Board*, número 1917, pp. 116-126. Washington, D.C.

ANEXOS

ANEXO I

Este anexo contém os gráficos da evolução do volume diário veicular dos postos P003, P011 e P012.

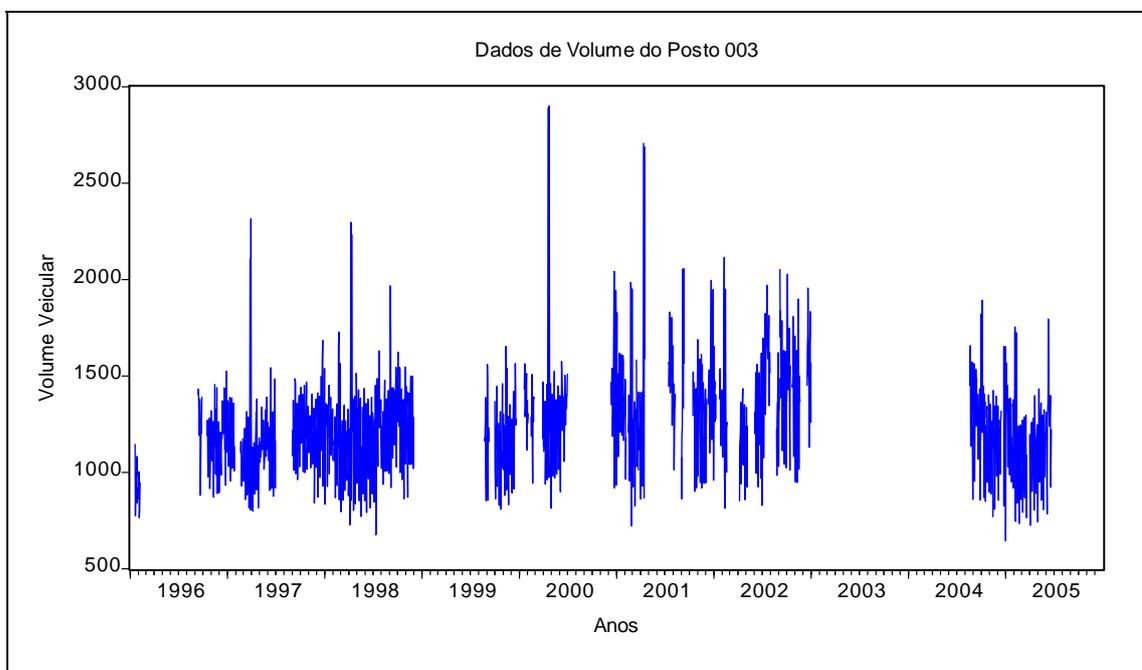


Figura I.1: Evolução do Volume Diário de Veículos no Posto P003

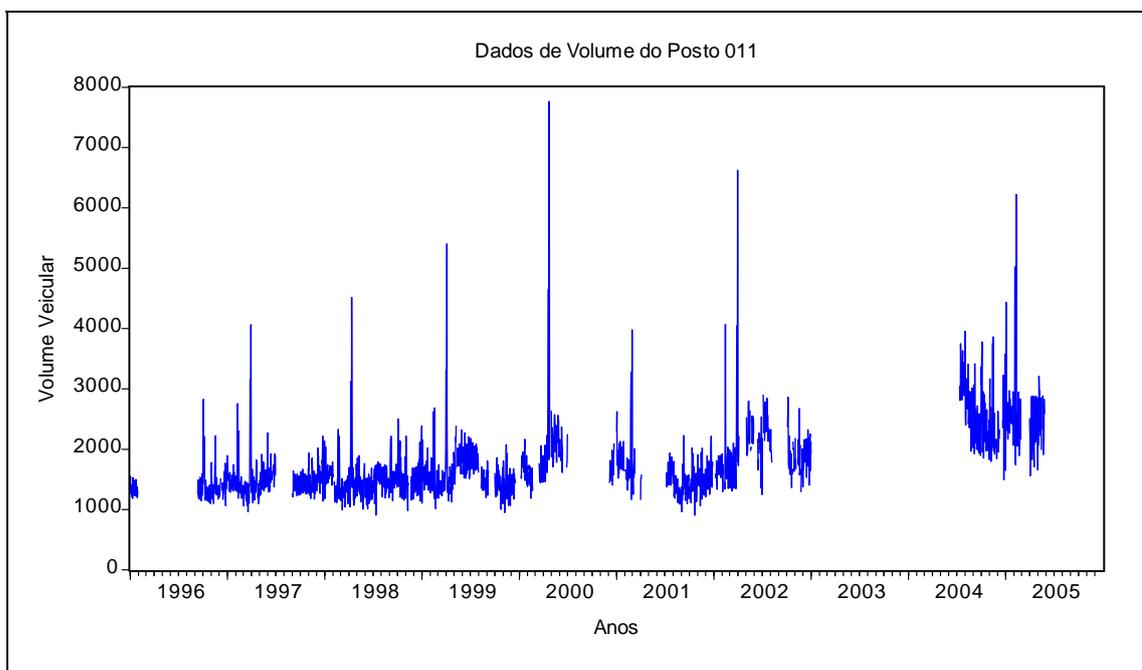


Figura I.2: Evolução do Volume Diário de Veículos no Posto P011

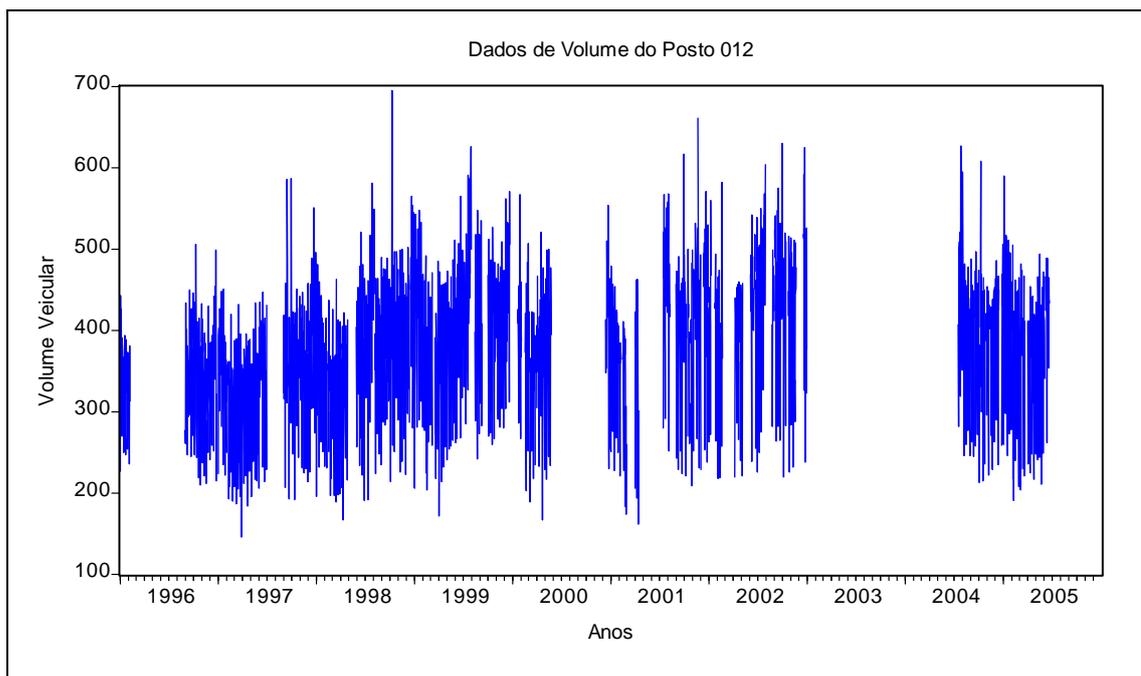


Figura I.3: Evolução do Volume Diário de Veículos no Posto P012

ANEXO II

Este anexo é composto por tabelas que mostram os resultados dos testes estatísticos de normalidade e estacionariedade para os postos P003, P011 e P012.

Tabela II.1: Testes de Normalidade e Estacionariedade para o Posto 003

Estatística	GL	Valor	p-valor
Jarque-Bera	2	393,520	< 0,0001
Box-Pierce	6	17322,357	< 0,0001
Ljung-Box	6	17349,465	< 0,0001
Box-Pierce	12	31719,005	< 0,0001
Ljung-Box	12	31793,738	< 0,0001

Tabela II.2: Testes de Normalidade e Estacionariedade para o Posto 011

Estatística	GL	Valor	p-valor
Jarque-Bera	2	179,874	< 0,0001
Box-Pierce	6	15312,064	< 0,0001
Ljung-Box	6	15336,084	< 0,0001
Box-Pierce	12	27786,435	< 0,0001
Ljung-Box	12	27852,084	< 0,0001

Tabela II.3: Testes de Normalidade e Estacionariedade para o Posto 012

Estatística	GL	Valor	p-valor
Jarque-Bera	2	407,457	< 0,0001
Box-Pierce	6	15369,518	< 0,0001
Ljung-Box	6	15393,472	< 0,0001
Box-Pierce	12	27854,669	< 0,0001
Ljung-Box	12	27919,821	< 0,0001

ANEXO III

Este anexo contém os autocorrelogramas dos postos P003, P011 e P012.

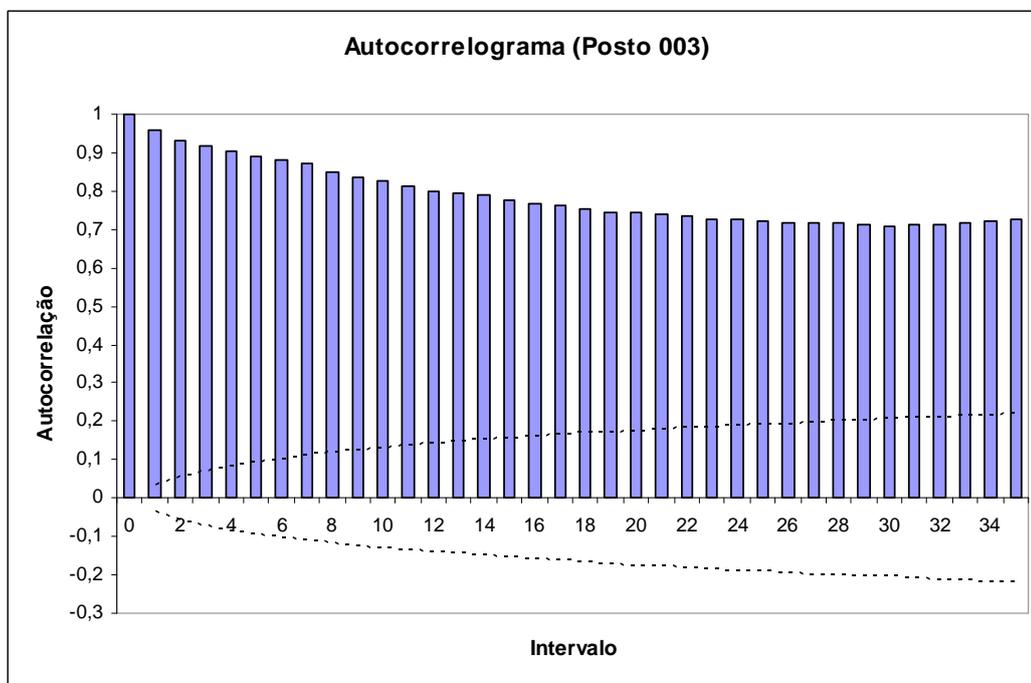


Figura III.1: Autocorrelograma do Posto 003

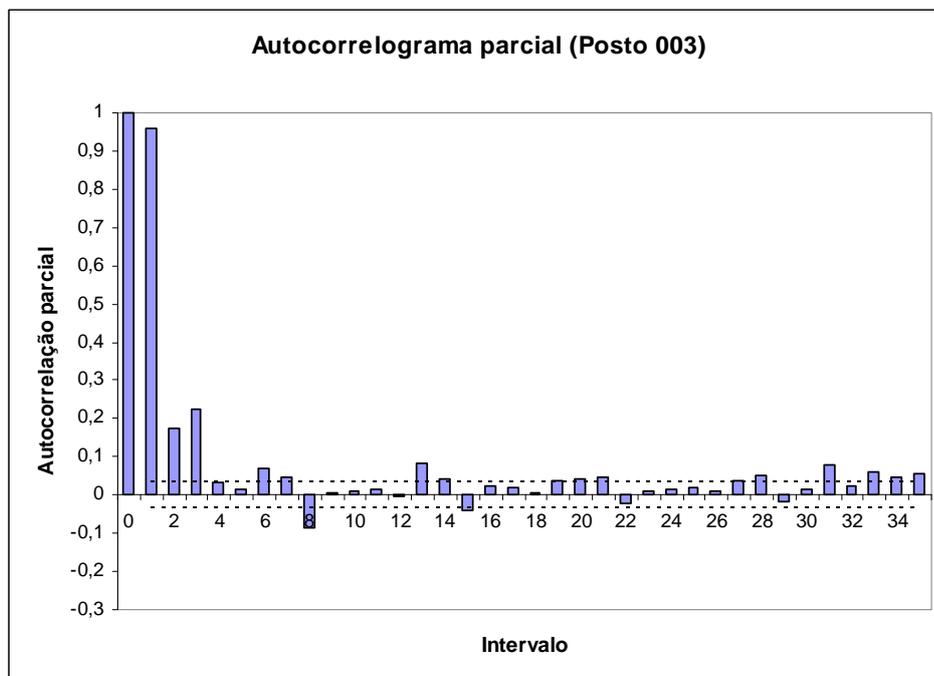


Figura III.2: Autocorrelograma Parcial do Posto 003

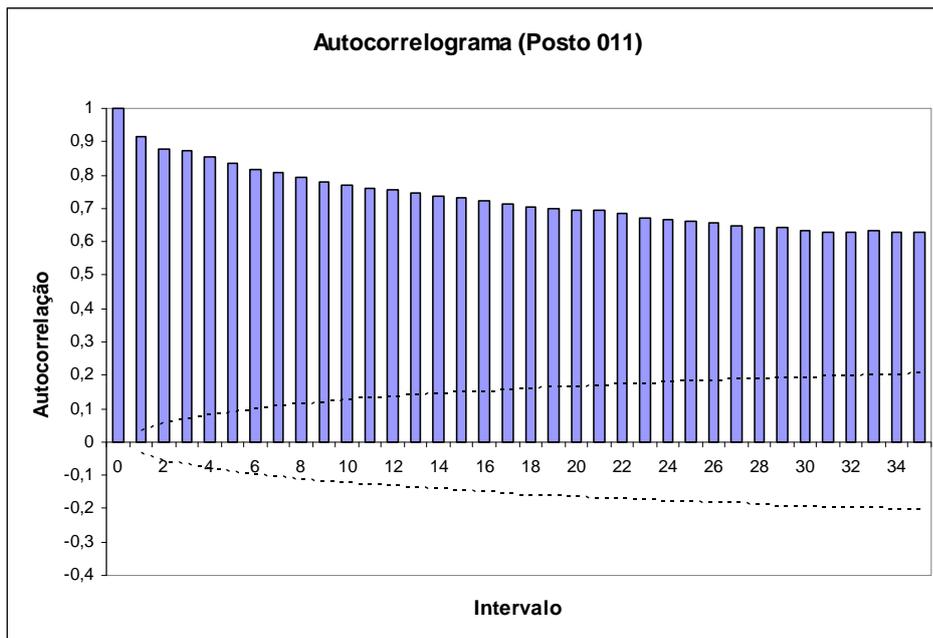


Figura III.3: Autocorrelograma do Posto 011

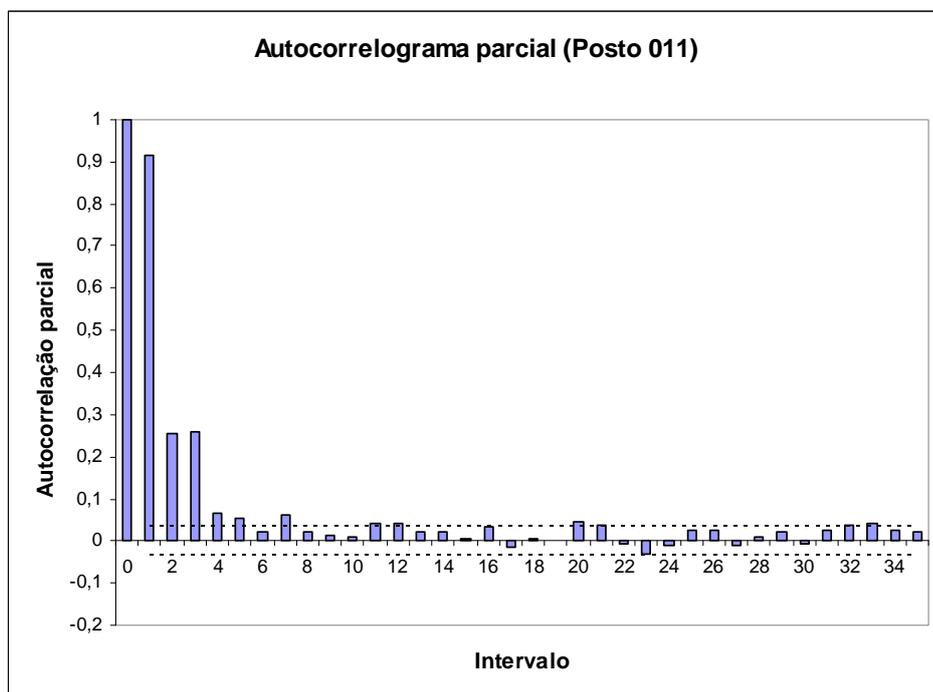


Figura III.4: Autocorrelograma Parcial do Posto 011

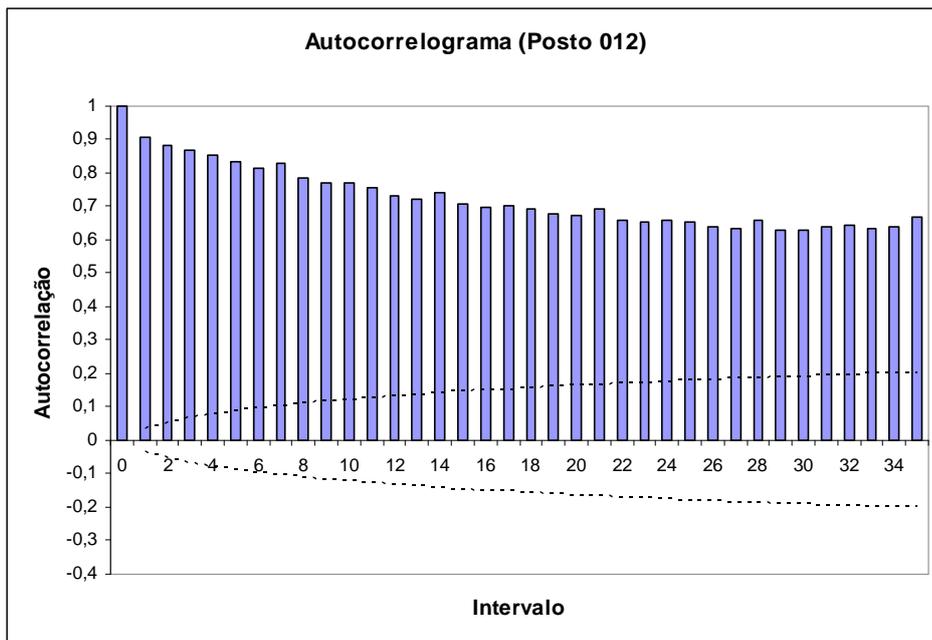


Figura III.5: Autocorrelograma do Posto 012

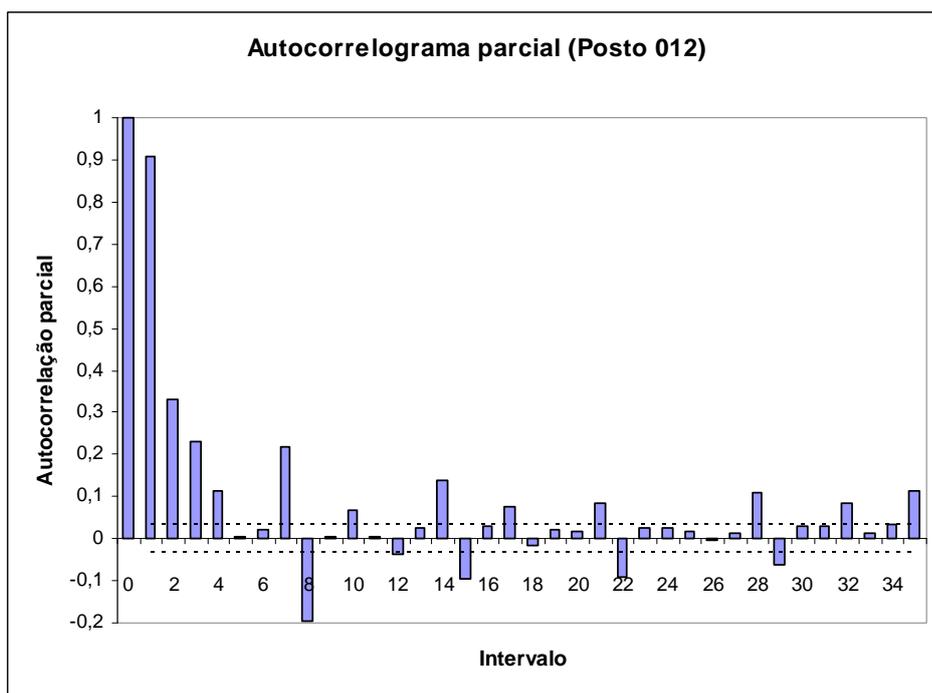


Figura III.6: Autocorrelograma Parcial do Posto 012