

ANÁLISE DA APLICAÇÃO DO MODELO ARIMA NA INSERÇÃO DE DADOS EM UMA SÉRIE HISTÓRICA DE VOLUME DE TRÁFEGO: ESTUDO DE CASO.

Antonia Fabiana Marques Almeida

Maria Elisabeth Pinheiro Moreira

Programa de Mestrado em Engenharia de Transportes - PETRAN

Universidade Federal do Ceará

RESUMO

Os programas de monitoramento de tráfego existentes nos órgãos rodoviários têm, como uma das atividades, a coleta de dados de volume veicular, para realização de estudos e planejamento de tráfego. Estes dados, porém, podem apresentar algumas falhas, ocasionadas da perda de dados, ou impossibilidade de coleta destes devido a problemas operacionais do método de contagem utilizado. Devido a estas falhas, os órgãos podem não utilizar estes dados, por não possuírem técnicas de complementação dos espaços vazios. Assim, este trabalho busca realizar a estimação de dados faltosos para um equipamento de uma rodovia estadual do Ceará e verificar a precisão obtida com a aplicação do modelo ARIMA de análise de séries temporais.

1. INTRODUÇÃO

O sistema de transportes desempenha um importante papel no desenvolvimento da economia de um país, uma vez que atende à crescente necessidade dos deslocamentos de bens e pessoas. No Estado do Ceará, como também nos outros estados brasileiros, estes deslocamentos, na sua maioria, ainda utilizam o sistema rodoviário, o que justifica a necessidade de um planejamento criterioso do setor, possibilitando uma aplicação racional dos recursos disponíveis (ASTEF, 1994). Para maior êxito deste planejamento, é necessário que os órgãos gestores sejam subsidiados com dados que representem a realidade das condições do tráfego, coletadas sob a forma de contagem de tráfego nas rodovias.

Pensando nisso, o Departamento de Edificações e Rodovias – DER, em 1996, iniciou a operação do programa de monitoramento contínuo do volume de tráfego nas rodovias estaduais do Ceará, compostos por 12 postos permanentes, que realizam contagens continuamente ao longo do ano.

Devido ao mau funcionamento e outros motivos (vandalismo, falta de manutenção, falta de coleta, etc.) os equipamentos contadores podem não registrar todos os dados de volume ao longo do ano, fazendo com que os parâmetros necessários aos estudos rodoviários sejam obtidos com base em dados incompletos (Sharma *et al.*, 2003).

Segundo Albright (1991) a preocupação no tratamento e análise de dados faltosos dos programas de monitoramento de tráfego existe desde os anos 30, e a prática de alguns órgãos rodoviários é a estimação dos valores perdidos ou inválidos, substituindo-os muitas vezes por valores históricos, sendo esta técnica denominada *imputation*.

A técnica de estimação de dados faltosos, não é recomendada segundo a AASHTO (1992), pois pode introduzir erros que não são quantificáveis. Porém, Sharma *et al.* (2003) afirmam que a estimação de dados faltosos não é proibitiva e que, para algumas análises avançadas pode ser realizada.

De acordo com pesquisa realizada por Zhong *et al.* (2002), os dados de tráfego geralmente apresentam quantidades significativas de valores faltosos. Em Alberta (Canadá), por exemplo, por 7 anos mais que a metade dos postos de contagem permanente tiveram valores faltosos, e

durante alguns anos estes percentuais variaram entre 70 e 90%. Segundo estudo realizado por Gomes e Moreira (2004), o percentual médio de dados faltosos para a série histórica coletada pelos equipamentos de monitoramento contínuo do volume de tráfego das rodovias estaduais do Ceará foi de 45%, percentual inferior aos apresentados para a cidade de Alberta.

Estudos realizados por Sharma *et al.* (2003), Williams *et al.* (1998), por exemplo, apresentam técnicas utilizadas na estimação dos valores faltosos, como a aplicação de fatores simples e a modelagem de séries temporais, com o uso de modelos estatísticos. Nestes estudos a utilização do modelo ARIMA apresentou melhor desempenho com relação aos demais modelos estatísticos utilizados.

Dessa forma, este trabalho propõe a estimação de dados faltosos de volume de tráfego, com o objetivo de verificar se os valores gerados podem ser utilizados na série histórica do programa de monitoramento de tráfego. Para tanto, será escolhido um posto de contagem de tráfego de uma rodovia estadual do Ceará, e os dados faltosos serão estimados com a utilização do modelo ARIMA (*Autoregressive Integrated Moving Average*) para análise e complementação.

2. REVISÃO BIBLIOGRÁFICA

Como as séries temporais apresentam peculiaridades, existem técnicas específicas para realizar estimações e previsões. As técnicas mais importantes e mais utilizadas são: Análise de Fourier, ou análise harmônica; Análise de Ondas (Wavelets); Amortecimento Exponencial (*Exponential Smoothing*); Processamento Temporal utilizando Redes Neurais Artificiais; ARIMA, etc. O modelo ARIMA, o qual será utilizado neste estudo, está brevemente descrito a seguir.

O ARIMA foi desenvolvido ao final dos anos 60, e Box e Jenkins (1976) o sistematizaram como um modelo que utiliza variações e regressões de dados estatísticos com o fim de encontrar padrões para uma estimação do futuro, podendo determinar quanto do passado se deve utilizar para prosseguir na previsão dos valores das séries (Ehlers, 2007).

O ARIMA necessita que sejam identificados os coeficientes e o número de regressões que serão utilizados e é bastante sensível à precisão com que se determinam seus coeficientes. É um modelo (p, d, q), onde: p se relaciona a Autoregressão, d está relacionado à Integração ou Diferenciação e q a Média Móvel, e é utilizado para decidir as estimações futuras que venham a ser explicadas pelos dados do passado e por variações independentes.

O uso do ARIMA tem sido difundido em variadas áreas de estudo de funções e previsões a fim de analisar as tendências dos valores e elementos de séries temporais. De acordo com Moretin e Toloi (2006), o modelo apresenta um grau de dificuldade considerável, uma vez que exige certa experiência de analistas e pesquisadores da área.

3. OBJETIVOS

Este trabalho tem como objetivo geral realizar análises comparativas das tendências dos dados de volume do programa de monitoramento de tráfego rodoviário do Estado do Ceará, fazendo o uso do modelo ARIMA, com base na série histórica de equipamento de contagem de tráfego de uma rodovia estadual do Ceará, a fim de aplicar e analisar esta metodologia na estimação dos dados faltosos. Para tanto, tem-se como objetivos específicos:

- a) Revisão bibliográfica acerca dos programas de monitoramento rodoviário existentes, dos modelos estatísticos de Análise de Séries Temporais e de estudos já realizados buscando analisar as tendências dos dados faltosos e do próprio modelo ARIMA;
- b) Realização de simulações em períodos de contagem sem dados faltosos, forçando a omissão de alguns dados para verificação dos resultados com a aplicação da metodologia ARIMA;
- c) Aplicação e validação do modelo, visando à complementação dos espaços vazios do banco de dados do programa de monitoramento de tráfego devido aos valores faltosos.

4. METODOLOGIA

Para elaboração deste estudo, as etapas a seguir apresentadas devem ser seguidas.

4.1. Obtenção dos Dados e Seleção do Posto de Contagem

Para aplicação do modelo proposto, serão coletados os dados de volume de tráfego dos 12 postos de contagem permanentes das rodovias estaduais do Ceará junto ao órgão com jurisdição sobre estas vias. Após a obtenção, os dados serão organizados e caracterizados para análise e aplicação do modelo.

4.2. Simulação da Aplicação do Modelo ARIMA

Para realização da simulação de utilização e análise da eficácia do ARIMA, será utilizada uma série histórica completa, ou seja, sem dados faltosos, e será forçada a omissão de alguns dados, para posterior aplicação do modelo, a fim de conhecer a geração de dados precisos e, assim, afirmar sua acurácia.

4.2. Geração do Modelo ARIMA

Os dados da série histórica do posto de contagem selecionado serão analisados, por meio do *software* STATISTICA (STATSOFT, 1991), buscando identificar qual o melhor modelo (p, d, q) para inserção dos dados faltosos na série, com o objetivo de complementá-la. Para tanto, será necessário identificar a sazonalidade existente e eliminá-la, tomando como suposição que dados de volume de tráfego são tendenciosos, no sentido que o fluxo corresponde ao período (férias, feriados, festividades, etc.). Assim, a validação se dará com o melhor modelo obtido, verificando os erros obtidos, na tentativa de minimizá-los, por isso é importante a utilização e análise dos modelos ARIMA.

4.3. Aplicação e Eficiência do Modelo ARIMA

Para medir a acurácia ou os erros percentuais obtidos pela estimação do modelo ARIMA proposto, será utilizada uma série histórica completa, sem dados faltosos, e retirado um percentual de dados de forma aleatória, comparando os valores estimados pelo modelo com os valores reais removidos. A quantidade de dados a serem removidos será definida de acordo com a construção de cenários de perdas de dados apresentadas pelos postos de contagem contínua variando entre percentuais de 5 a 50%.

5. CONSIDERAÇÕES FINAIS

Com a conclusão desta pesquisa, o órgão rodoviário irá dispor de uma metodologia para análise, tratamento e inserção de dados faltosos de contagens de tráfego dos postos permanentes nas rodovias estaduais do Ceará, de forma a ter um banco de dados completo com os volumes veiculares rodoviários, auxiliando os processos de planejamento e tomada de decisão embasados nestas informações, para melhoria da malha viária.

Como contribuição científica, este trabalho busca analisar a eficácia do modelo ARIMA na estimação de dados faltosos para séries temporais, dando suporte a pesquisadores no conhecimento desta técnica de análise.

REFERÊNCIAS BIBLIOGRÁFICAS

- AASHTO (1992) *Guidelines for Traffic Data Programs*. American Association of State Highway and Transportation Officials. Washington, D.C.
- ALBRIGHT, D. (1991) History of Estimating and Evaluating Annual Traffic Volume Statistics. *Transportation Research Record 1305*, Transportation Research Board, Washington, D.C, pp. 103-107.
- ASTEF (1994) *Plano de Contagem de Tráfego – Relatório Final*. Associação Técnico-Científico Eng. Paulo Froton, Universidade Federal do Ceará, Fortaleza, Ceará.
- 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.
- EHLERS, R. S. (2007) *Análise de Séries Temporais*. Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná.
- GOMES, M. J. T. L.; MOREIRA, M. E. P. (2004) Consolidação da Base de Dados de um Programa de Monitoramento Contínuo do Volume de Tráfego das Rodovias Estaduais do Ceará. *Anais do XVIII Congresso de Ensino e Pesquisa em Transportes. ANPET - Associação Nacional de Pesquisa e Ensino em Transportes*.
- MORETTIN, P. A.; TOLOI, C. M. C. (2006) *Análise de Séries Temporais*. 2ª edição revisada e ampliada. Ed. Edgard Blücher. São Paulo.
- SHARMA, S.; LINGRAS, P. E ZHONG, M. (2003) Effect of Missing Value Imputations on Traffic Parameters Estimations from Permanent Traffic Counts. *Transportation Research Board*, The 82th Annual Meeting, Washington, D.C., USA.
- STATSOFT (1991) “STATISTICA – Reference for Statistical Procedures”, Statsoft Inc, Tulsa, OK.
- WILLIAMS, B. M.; DURVASULA P. K. E BROWN D. E. (1998) “*Urban Freeway Traffic Flow Prediction: Application of Seasonal ARIMA and Exponential Smoothing Models*”. *Transportation Research Record* #1644 pp. 132-141. National Academy of Sciences.
- ZHONG, M.; LINGRAS, P.; E SHARMA, S. (2002) Updating Missing Values of Traffic Counts: Factor Approaches, Time Series Analysis versus Genetically Designed Regression and Neural Network Models. *Transportation Research Part C*, United Kingdom, 2002.

Antonia Fabiana Marques Almeida (fabiana@det.ufc.br)

Maria Elisabeth Pinheiro Moreira (beth@det.ufc.br)

Departamento de Engenharia de Transportes

Centro de Tecnologia - Universidade Federal do Ceará Campus do Pici - Bloco 703 - 60455-760 Fortaleza/CE

Fone/Fax: (85) 3366-9488