

CONSOLIDAÇÃO DE DADOS DE FISCALIZAÇÃO ELETRÔNICA PARA OBTENÇÃO DO VOLUME DE TRÁFEGO E DO FLUXO ORIGEM-DESTINO VEICULAR

Joana Maia Fernandes Barroso
João Lucas Albuquerque Oliveira
Francisco Moraes de Oliveira Neto
Universidade Federal do Ceará
Departamento de Engenharia de Transportes

RESUMO

O conhecimento da dinâmica dos deslocamentos viários no meio urbano é essencial para o planejamento operacional e estratégico dos sistemas de transportes. Recentemente, é crescente a disponibilidade de dados massivos de tráfego urbano provenientes de sistemas de identificação e coleta automática de veículos, permitindo uma compreensão da variabilidade dia-a-dia do tráfego veicular em sistemas viários de grandes cidades. Porém, para garantir a confiabilidade das análises, é essencial a realização de uma consolidação prévia desses dados. Este trabalho apresenta um método de consolidação de dados de Fiscalização Eletrônica para a obtenção de séries temporais de volumes de tráfego e de fluxos origem-destino para diferentes períodos do dia. O método proposto foi aplicado para gerar séries de tráfego a partir dos dados do sistema de Fiscalização Eletrônica na cidade de Fortaleza-CE, contribuindo para identificar dados suspeitos e atípicos, definir períodos representativos do tráfego veicular e viagens entre zonas do meio urbano.

ABSTRACT

The knowledge of the dynamic of trip pattern in the urban environment is essential for the operational and strategic planning of transportation systems. Recently, the increasing availability of massive traffic data from automated vehicle identification and collection systems can allow an understanding of the day-to-day variability of traffic flows in large urban network systems. However, to enhance the data quality for analysis, it is essential to carry out a previous data treatment. This work presents a method for treatment of data from Traffic Surveillance systems to obtain time series of traffic volumes and origin-destination flows (OD flows) for different periods of a typical day. The proposed method was applied to generate traffic series from the Traffic Surveillance system of Fortaleza city, contributing to identify suspicious and atypical data, to define representative periods of vehicular traffic and to estimate series of OD flows.

1. INTRODUÇÃO

O padrão de deslocamento viário de uma cidade pode ser representado principalmente por duas variáveis: os fluxos origem-destino (OD), que indicam a quantidade de viagens realizadas entre zonas de uma área de estudo, e o volume de tráfego, que indica a demanda em nós e arcos da rede de transportes. Tanto a distribuição dos fluxos OD como a magnitude dos volumes representam informações básicas para o planejamento e projeto de transportes, bem como para o gerenciamento e controle do tráfego (Cremer e Keller, 1987). Os fluxos OD são resultado das decisões de viagem de uma população com o objetivo de realizar atividades localizadas no espaço urbano. Em contraste com os dados de volume que são facilmente coletados por dispositivos de contagem, os fluxos OD demandam muito esforço para serem medidos diretamente, sendo necessárias entrevistas individuais ou levantamentos de placas. Diante disto, houve um esforço considerável no passado para desenvolver métodos que forneçam estimativas para os fluxos OD a partir de contagens de tráfego nos arcos de uma rede (Pitombeira Neto e Loureiro, 2016; Pitombeira Neto *et al.*, 2018). Já a variável volume de tráfego surge como resultado dos fluxos OD e da alocação desses fluxos na rede, e pode ser definida como o total de veículos que trafega em uma seção da via em uma determinada faixa ou direção durante um intervalo específico de tempo (Roess e McShane, 2004).

Segundo Cascetta (2009), o conhecimento do padrão de deslocamento dos usuários é essencial para a formulação e implementação de modelos que representem os elementos que compõem o sistema de transportes, e que auxiliem na tomada de decisão quanto a intervenções na oferta do sistema. Geralmente, nos estudos de demanda, é comum se observar a variação temporal do tráfego dia-a-dia ou ao longo do dia. Atualmente, existem alguns tipos de dados que podem ser utilizados para se obter informações sobre os padrões de deslocamento de uma população, como dados de publicações em redes sociais ou de telefonia móvel, que podem gerar informações sobre o deslocamento dos usuários desses serviços, dados de sistemas GPS (Global Positioning System) ou de leituras de placas veiculares, que fornecem informações sobre o deslocamento de veículos na malha viária, e dados de bilhetagem eletrônica, que podem auxiliar na compreensão dos deslocamentos dos usuários de transporte público. Na cidade de Fortaleza/CE existem aproximadamente trezentos equipamentos de fiscalização eletrônica (FE) distribuídos na malha viária, localizados principalmente nas vias arteriais de maior fluxo veicular. Esses equipamentos registram a passagem de cada veículo e realizam a leitura e registro das placas de uma proporção dos veículos detectados. A partir desses dados, é possível extrair dados do volume veicular registrado por cada um dos equipamentos, assim como associar leituras de placas entre equipamentos de regiões distintas da cidade e obter informações acerca dos fluxos OD entre elas.

Apesar de sua grande disponibilidade, os dados de FE possuem algumas limitações para a utilização em estudos de demanda. A concentração e a distribuição dos equipamentos em vias arteriais da cidade faz com que não seja possível observar muitas das rotas utilizadas pelos usuários. Além disso, a informação sobre viagens obtidas a partir da associação de placas representa apenas um fluxo de passagem, não garantindo que seja de fato um fluxo OD. Ainda devem ser consideradas possíveis falhas dos equipamentos durante a operação do sistema de coleta de dados, tanto em relação à detecção dos veículos, quanto à leitura de placas. Tanto nas aplicações que analisam a variabilidade do tráfego dia-a-dia, quanto ao longo do dia, é necessária a definição de períodos de pico e entropico, ou seja, a separação do dia em períodos nos quais seja possível considerar o fluxo veicular relativamente estável. A partir disso, este trabalho tem como objetivo principal propor uma metodologia de consolidação de dados provenientes de sistemas de FE para a obtenção de séries temporais de volumes de tráfego e fluxos OD dia-a-dia para um longo período de tempo. Os objetivos específicos são: *i*) tratar dados originais de FE para extrair o volume de tráfego e os fluxos OD, definindo períodos de tempo em que o tráfego se mantém estável e eliminando o efeito das possíveis falhas nos sensores de coleta e na captura e leitura de placas ao longo do tempo; *ii*) avaliar se as distribuições das variáveis obtidas se adequam ao que se afirma na literatura.

2. REVISÃO

A estratégia de coleta automática de dados está cada vez mais presente nos sistemas de transportes no mundo todo, tendo sido inicialmente projetada com objetivos de navegação e localização de veículos, controle de embarque e arrecadação no transporte público e controle operacional do tráfego urbano. Alguns exemplos de estratégias desse tipo são os sistemas de localização automática de veículos, a contagem automática de passageiros, sistemas de bilhetagem eletrônica, e sistemas de contagem de tráfego e leitura de placas. A coleta automática de dados de tráfego também pode subsidiar sistemas de informação em tempo real para usuários, assim como auxiliar pesquisadores para a melhor compreensão dos padrões de deslocamento das populações nas cidades. Zhao *et al.* (2010) destacam o potencial do uso de

dados provenientes de sistemas de coleta automática de dados no planejamento de transportes, apesar de esses sistemas terem sido projetados para outras funções bastante específicas. Para ilustrar esse potencial, os autores apresentam um método para inferir uma matriz OD do sistema de transporte ferroviário utilizando dados coletados de bilhetagem eletrônica da cidade de Chicago.

O sistema de FE instalado na cidade de Fortaleza é constituído por sensores indutivos instalados no pavimento, responsáveis pela detecção dos veículos, e por um sistema de câmeras juntamente com um software OCR (*Optical Character Recognition*), responsável pela captura e identificação das placas dos veículos detectados. Este sistema de FE pode ser considerado um sistema de coleta automática de dados, assim como os sistemas de bilhetagem eletrônica segundo as propriedades destacadas por Chu *et al.* (2010), como a estratégia de coleta de dados passiva, em que os dados são gerados automaticamente sem a necessidade de informações declaradas ou relatadas por pessoas entrevistadas, a pontualidade e período de tempo de coleta prolongado, sendo contínua enquanto o sistema estiver em operação, a integração com o planejamento operacional, gerando dados para auxiliarem intervenções, e a resolução desagregada dos dados, em que cada transação é um registro individual de um veículo com detalhes operacionais.

A extração de informações de tráfego veicular como velocidade média, tempo de viagem, volume e fluxos OD a partir de dados provenientes sistemas de coleta automática tem sido explorada na literatura, como em Castillo *et al.* (2008) que utilizam dados de leituras de placas, juntamente com contagens de volume de tráfego em arcos da rede para reconstruir uma matriz OD e o fluxo na rede, e em Bertini *et al.* (2005) que utilizam dados de leituras de placas para previsão do tempo de viagem de veículos em um corredor e comparam os resultados com os obtidos utilizando GPS. Porém, na maioria dos trabalhos, o foco principal é na análise dessas variáveis, sem muito detalhamento da etapa anterior de tratamento e preparação dos dados. Por outro lado, Oliveira e Loureiro (2006) apresentam a etapa de consolidação de dados do sistema de coleta automática presente na cidade de Fortaleza-CE, com o objetivo de excluir da amostra dias com falhas, e dias atípicos como finais de semana e feriados, assim como Stathopoulos e Karlaftis (2001) definem seis períodos de tráfego distintos, como resultado de uma investigação prévia dos padrões de viagem da cidade de Atenas, Grécia.

A etapa de consolidação dos dados utilizados em análises de tráfego e mobilidade é essencial para que se possa confiar na capacidade dos dados de refletirem a realidade estudada. Nesta etapa busca-se eliminar qualquer viés, seja devido a falhas na coleta automática ou quando só é possível coletar uma amostra referente a categorias específicas da população. Caso não seja realizada uma análise prévia dos dados, os resultados das análises das variáveis de interesse podem ser prejudicados, como apontam Cheng *et al.* (2012), que analisaram a autocorrelação espaço-temporal de dados de tráfego, utilizando dados provenientes de sensores localizados na malha viária da cidade de Londres, Reino Unido. Para examinar como a autocorrelação varia ao longo do dia, os autores assumiram três períodos distintos, pico da manhã, pico da tarde e entropico, definidos pelo órgão governamental TfL (*Transport for London*). O estudo constatou a presença de dependência sazonal no período de entropico, o que não era esperado, dado que esse período deveria capturar um estado que separa os dois picos, caracterizado pelo retorno às condições de fluxo livre, e, a partir disso concluíram que a divisão arbitrária dos

dados em três períodos de tempo baseado em definições aceitas não é suficiente para isolar estados específicos de tráfego na rede.

Os resultados dos procedimentos de tratamento de grandes volumes de dados de transportes provenientes de sistemas de coleta automática são, não apenas essenciais para a realização de análises confiáveis, como também o primeiro passo para a compreensão da dinâmica do tráfego em um ambiente urbano. Ou seja, além dos procedimentos de limpeza e organização, a etapa de consolidação também inclui uma análise descritiva preliminar dos dados com o objetivo de identificar padrões que irão auxiliar no levantamento de hipóteses a serem testadas sobre o fenômeno estudado, possibilitando assim, a obtenção de resultados que melhor representem a realidade.

3. MÉTODO

Os dados de FE utilizados como referência para a construção da metodologia proposta foram coletados na cidade de Fortaleza-CE, onde existem três empresas responsáveis pelos equipamentos de FE. Os arquivos, no formato CSV (*Comma-Separated Values*), de dados coletados por cada uma das empresas são fornecidos da Autarquia Municipal de Trânsito e Cidadania (AMC). Os arquivos de um dia contêm registro de cada veículo detectado com as seguintes informações: código de identificação do equipamento, data hora da detecção; faixa da via na qual o veículo foi detectado, a velocidade regulamentada da via e a velocidade instantânea do veículo, o tamanho do veículo detectado em metros, a classificação do tipo de veículo e a leitura da placa criptografada do veículo detectado.

A seguir serão descritas as etapas do método de consolidação de dados de FE, conforme Figura 1. A etapa de pré-processamento consiste na organização inicial dos dados em arquivos únicos por empresa e por dia. Os dados são então agregados em intervalos de 5 minutos, com o objetivo de observar um período pequeno o suficiente para que se possam observar variações com maiores detalhes dentro do intervalo horário, porém não tão pequeno ao ponto de se tornar comum obter detecções de volume iguais à zero.

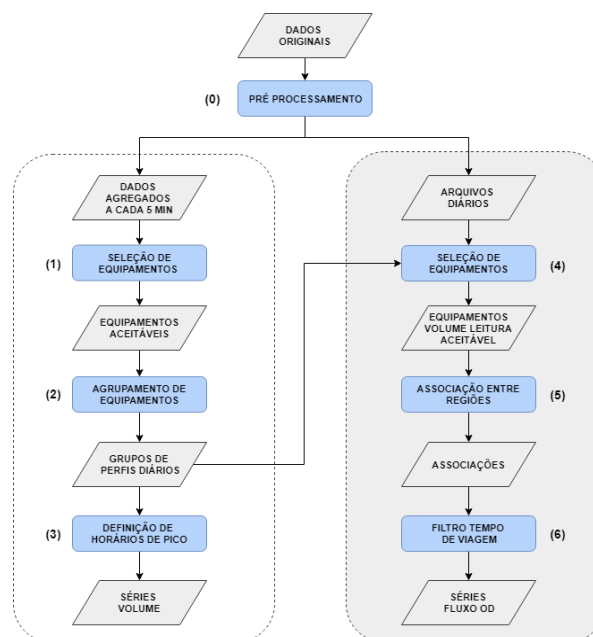


Figura 1: Metodologia proposta para consolidação dos dados

3.1. Extração do Volume de Tráfego

3.1.1. Seleção de Equipamentos

A etapa de seleção dos equipamentos (1) tem como objetivo garantir a seleção de equipamentos que tenham um funcionamento considerado aceitável para a obtenção dos dados de interesse durante o período de análise. Inicialmente são selecionados os equipamentos que tenham funcionado em todos os meses do período, ou seja, sem nenhum mês faltante na amostra. Em seguida os equipamentos são filtrados com base na proporção de intervalos de 5 minutos, no período entre 5:00h e 22:00h, em que houve leituras de volume diferentes de zero, ou seja, quanto mais próximo de 1, mais confiável é considerado o equipamento. Considerou-se que no período do dia analisado não se observaria volumes de 5 minutos iguais a zero, dado que grande parte dos equipamentos de fiscalização se localiza em vias arteriais da cidade. Definiu-se os seguintes critérios de seleção: proporção de zeros aceitável em um dia, e total de dias aceitáveis de um equipamento. Para definir o primeiro, realizou-se uma análise de sensibilidade buscando um compromisso entre a ponderação do critério e a representatividade da amostra. Em relação ao segundo critério, foi estabelecido que um equipamento seria aceitável caso apresentasse um total de dias aceitos pelo critério anterior, maior ou igual a 90% do total de dias úteis presentes no período da amostra.

3.1.2. Agrupamento de Equipamentos

A etapa seguinte se refere ao agrupamento dos equipamentos (2) de acordo com seu perfil relativo de fluxo horário médio. Weijermars (2007) utilizou técnicas de clusterização para identificar perfis de volume de tráfego, o que pode ser útil para a identificação dos horários de pico de acordo com cada perfil, considerando-se que tráfego pode variar de acordo com a região da cidade e do sentido de deslocamento dos usuários. Para realizar o agrupamento, utilizou-se o algoritmo *k-means*, que define grupos com base no número de agrupamentos pré-definido e numa medida de similaridade entre cada elemento a ser agrupado. Considerou-se como medida de similaridade a distância euclidiana entre os volumes relativos no mesmo intervalo de 5 minutos ao longo do dia, logo, gerando um vetor de distâncias de tamanho igual a 288, referente à quantidade de intervalos de 5 minutos em um dia. Para definir o número de *clusters*, partiu-se inicialmente de um índice de critério interno, o *average silhouette width* que busca maximizar a separabilidade e compacidade dos grupos. Em seguida, para a validação desse valor, foram avaliados valores adjacentes de agrupamentos, buscando aliar a qualidade dos *clusters* obtidos com a melhor representação do fenômeno.

3.1.3. Definição de Horário de Pico

A definição dos horários de pico (3) é realizada para cada perfil identificado na etapa anterior, visando definir períodos do dia em que o volume veicular pode ser considerado estável, incluindo também a identificação de um horário de entrepico. Para tanto, utilizou-se o algoritmo de clusterização *Density-based spatial clustering of applications with noise* (DBSCAN), proposto por Ester *et al.* (1996), que é adequado para identificar clusters arbitrários de diferentes tamanhos, e identificar elementos que não pertencem a nenhum cluster, chamados ruído, sem a necessidade de fornecer informações preliminares sobre os grupos.

3.2. Extração do Fluxo OD

3.2.1. Seleção de Equipamentos

A etapa seguinte (4) diz respeito à seleção de equipamentos adequados para a extração da variável fluxo OD a partir da amostra de equipamentos previamente selecionada para a

extração do volume. Os sistemas de fiscalização eletrônica possuem a limitação de não registrarem todas as placas de veículos detectados, além de não fazerem toda a cobertura da malha viária. Por conta disso, esta etapa consiste na análise de cada um dos equipamentos para verificar a proporção de leituras ao longo do período de interesse. Como se pretende capturar a variabilidade do fenômeno através desses dados, a proporção de leituras deve variar o mínimo possível para que se possa amenizar o efeito da variação do desempenho dos equipamentos na amostra. Para esta análise, foram definidos períodos mais longos de observação que incluem os períodos de pico do tráfego, definido na Etapa (3), com início do período antes e final depois do período de tráfego estável, supondo-se que estes períodos contemplam a maior parte das viagens no período da manhã, da tarde e no entropico. Estes períodos foram definidos comparando os perfis de fluxo dos equipamentos entre regiões distintas. Definiu-se como critério de seleção dos equipamentos o desvio padrão da proporção amostral de leituras em cada período de análise, sendo considerando um limite máximo através de uma análise de sensibilidade, buscando equilibrar o compromisso existente entre confiabilidade dos dados (efeito mínimo da variação do desempenho dos equipamentos) e viabilidade da análise (quantidade de equipamentos).

3.2.2. Associação Entre Regiões

A etapa de associação entre equipamentos (5) consiste na identificação de deslocamentos realizados por veículos entre regiões da área de estudo nos períodos de pico e entropico. Inicialmente, é definido um zoneamento para a área de estudo, considerando-se a densidade de equipamentos em cada uma, assim como suas características de uso do solo e atividades. A partir do zoneamento definido, são contabilizados os deslocamentos entre equipamentos de fiscalização eletrônica presentes em regiões distintas da cidade através da associação de leituras de placas em um mesmo dia e dentro dos períodos do dia adotadas. Os veículos registrados em mais de dois equipamentos podem gerar registros a mais de deslocamentos, o que é solucionado mantendo apenas o deslocamento com maior tempo de viagem. A grande questão nesta etapa da consolidação é como representar o fluxo OD entre regiões da cidade através do fluxo de passagem entre equipamentos. Dado que a proposta deste trabalho é obter uma série temporal indicando como esse fluxo varia dia-a-dia, assume-se a premissa de que essa proporção, tanto do total de veículos que trafegam na malha viária, quanto da leitura de placas se mantém constante ao longo dos dias. Em segundo lugar, assume-se que o trecho da viagem contabilizada pela associação entre o primeiro e o último registro da placa dentro do período analisado de fato teve origem na região da primeira detecção e destino na última.

3.2.3. Filtro Tempo de Viagem

A última etapa da consolidação dos dados (6) diz respeito à análise da distribuição dos tempos de viagem das associações obtidas na etapa anterior. Inicialmente, são removidas da amostra observações com tempo de viagem menor que 30 minutos, com o objetivo de eliminar associações entre equipamentos muito próximos que tendem a gerar muitas observações, o que pode levar a um aumento nas viagens por conta desses pares de equipamentos, viesando assim a amostra. Em seguida os tempos de viagem observados são comparados com os tempos de viagem obtidos pela plataforma do Open Street Map (OSM) a partir do pacote 'osrm' desenvolvido por Giraud (2018) que faz parte do software R. Assim, os tempos de viagem entre os equipamentos analisados foram comparados com os tempos de viagem observados e filtrados de acordo com o critério da seguinte equação: $0,85 \times t_{osm} < \Delta t < 3 \times t_{osm}$, em que Δt é a diferença de tempo entre associações, e t_{osm} é o tempo obtido pelo 'osm'. Esta filtragem tem como objetivo eliminar observações que sejam casos de associações geradas

por leitura incorreta de placas, como tempos de viagem muito pequenos entre equipamentos distantes e vice-versa. O limite inferior de 85% do tempo fornecido pelo OSM se deve à aceitação da possibilidade de falhas de sincronização no sistema de fiscalização eletrônica, o que pode levar a uma defasagem entre os horários de equipamentos diferentes, gerando assim tempos de viagem menores do que o que de fato ocorreu.

3.3. Geração das Séries Temporais

Após definidos o período de pico e os deslocamentos com tempos de viagem aceitáveis, é possível gerar as séries de volume de tráfego para cada equipamento e de fluxo OD para cada par de regiões. Nessa etapa, realiza-se uma análise de *outliers* que não foram detectados pelos critérios anteriores, possivelmente devidos a acidentes, condições climáticas, entre outros motivos. O critério adotado foi de considerar um valor extremo, ou *outlier*, uma observação fora do seguinte intervalo: $[\tilde{x} - 3s, \tilde{x} + 3s]$, em que \tilde{x} e s são a mediana e o desvio padrão amostral, para uma amostra obtida utilizando uma janela temporal de 20 dias. Esta janela temporal foi definida com o intuito considerar a possível sazonalidade existente nos dados.

Para analisar qual distribuição de probabilidade pode representar a variação das séries temporais obtidas, as distribuições amostrais considerando uma amostra de meses típicos do ano foram confrontadas com a distribuição Normal. A suposição de que o modelo normal pode representar a distribuição do volume veicular e dos fluxos ODs é defendida por Oliveira e Loureiro (2006) e Pitombeira Neto (2015), respectivamente. De fato, acredita-se que os valores das variáveis de tráfego oscilam em torno de uma medida central, já que as séries serão geradas para períodos estáveis do tráfego ao longo do dia, ou de tráfego pendular, em que o padrão de fluxo se repete a cada dia. Para testar a hipótese de normalidade das duas variáveis de tráfego, utilizou-se o teste Shapiro-Wilk sem contabilizar os *outliers* identificados, além dos meses atípicos de janeiro, julho e dezembro.

4. ESTUDO DE CASO: FORTALEZA-CE

Nesta seção serão apresentados e discutidos os resultados obtidos a partir da aplicação do método de consolidação proposto para os dados de FE do ano de 2017 na cidade de Fortaleza-CE.

4.1. Extração do Volume de Tráfego

Inicialmente, estavam disponíveis dados de 358 equipamentos de FE ativos no ano de 2017 em Fortaleza. Após a primeira filtragem, restaram 271 equipamentos que possuíam dados de detecções de todos os meses presentes na amostra. Por fim, o critério para a filtragem de equipamentos em função da proporção de intervalos de 5 minutos com detecção diferente de zero foi definido a partir de uma análise de sensibilidade da variação do número de dias aceitáveis para todos os equipamentos, em função proporção de intervalos aceitáveis para um dia de detecção. Variando a proporção de intervalos de 5 minutos diferente zero de 60% até 100%, observou-se que a proporção de 90% correspondia a uma situação em que pelo menos 200, aproximadamente, do total equipamentos eram considerados aceitáveis, que corresponde a uma situação de equilíbrio entre o tamanho da amostra desejável e a qualidade dos dados. Com a indicação dos dias aceitáveis para cada equipamento, foi aplicado o segundo critério referente à quantidade mínima de dias aceitáveis, em que foram excluídos equipamentos com menos de 90% do total de dias úteis presentes no período da amostra, resultando em uma amostra de 144 equipamentos que apresentaram um bom funcionamento durante o período analisado, o que representa apenas cerca de 40% do total inicial de dados para 2017.

Com os equipamentos definidos, realizou-se o agrupamento dos perfis médios relativos de volume dos equipamentos utilizando o *k-means*. O indicador de critério interno, *average silhouette width*, retornou o valor 2 como número ótimo de *clusters*. Em seguida, para validar este valor, foram gerados agrupamentos com 2, 3 e 4 *clusters*. A partir disso foi verificado que, para a melhor compreensão do fenômeno, o número de 3 *clusters* seria mais adequado, representando melhor as diferentes dinâmicas de tráfego existentes na cidade. Na Figura 2, são apresentados os agrupamentos obtidos após aplicação do algoritmo. É possível perceber que no perfil 1 há um pico mais acentuado no período da noite, enquanto no 3 esse período se transfere para manhã de forma mais intensa. Observa-se também outros dois períodos de pico ao longo do dia, porém com intensidade menor. Já no perfil 2, há a ocorrência de dois períodos de pico bem definidos e de mesma intensidade e, entre eles, há uma espécie de platô com pequenas variações.

A disposição espacial dos equipamentos selecionados, o grupo ao qual eles foram atribuídos, bem como o sentido de tráfego detectado pelo equipamento foram avaliados para cada perfil. Os equipamentos do perfil 3 estão majoritariamente na direção periferia-centro, enquanto os equipamentos pertencentes ao perfil 1 estão normalmente no sentido inverso, revelando a característica predominantemente pendular das viagens, indo em direção ao centro no início do dia e retornando para residência no final do dia. Os equipamentos do perfil 2 indicam locais onde geralmente o tráfego não sofre alteração significativa em função da hora do dia, e estão normalmente em vias importantes que ligam bairros e até municípios, o que pode justificar a ocorrência de dois picos bem definidos, além do platô entre os dois.

A partir dos perfis identificados, utilizou-se o algoritmo DBSCAN para estimar os diversos horários (agrupamentos) em que o volume estivesse mais estável. O uso desse algoritmo em vez do *k-means* se deu em razão das possibilidades de identificação de um grupo de ruído e não definição *a priori* dos agrupamentos, uma vez que o intuito é separar períodos do dia. Além disso, O DBSCAN é mais adequado para gerar grupos cujos elementos possuem intensidade similar, enquanto o *k-means* gera grupos cujos elementos apresentam tendências de variação similares. Para todos os perfis, os dados de entrada para o algoritmo foram $MinPts = 12$, como valor mínimo de intervalos de 5 minutos, na tentativa de agrupar intervalos maiores ou iguais a uma hora e $\epsilon = 80$, como o raio de abrangência de cada ponto, obtido a partir da observação do gráfico de distância acumulada dos 12 ($MinPts$ definido) vizinhos mais próximos. Os agrupamentos identificados dentro do dia, para cada perfil, estão apresentados na Figura 2. As observações atribuídas ao cluster 0 são os ruídos e estão basicamente em todos os trechos com alta taxa de variação do volume veicular, o que implica na não estabilidade do volume de tráfego nesse período.

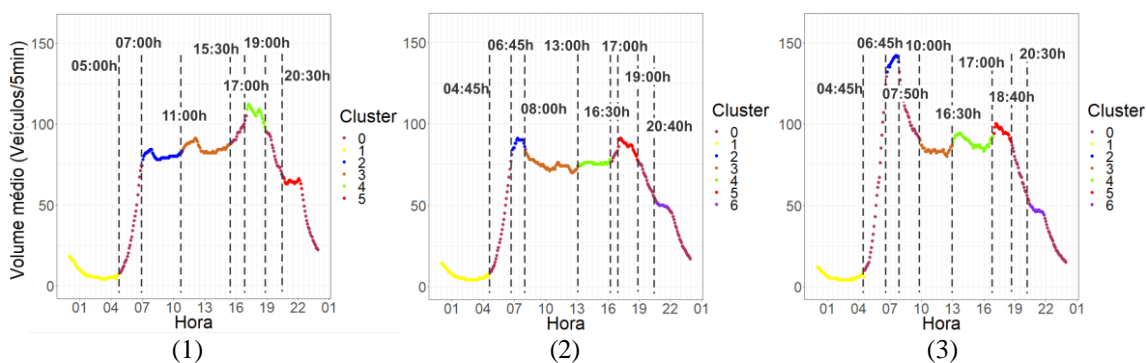


Figura 2: Perfis médios identificados e horários de pico

No perfil 1, observa-se que não há um período acentuado de pico pela manhã. Já no final da manhã e no período da tarde, o volume é um pouco maior e há um agrupamento definido, entre 11:00h e 15:30h. No horário entre 17:00h e 19:00h, nota-se a definição de um horário de volume mais estável para o período da tarde/noite. Para o perfil 2, observa-se mais claramente a definição de dois horários de pico, um pela manhã, das 06:45h às 08:00h e outro no período da tarde adentrando a noite, de 17:00h às 19:00h. No intervalo entre os picos, há a ocorrência de dois agrupamentos, um que se estende mais no período da manhã (08:00h às 13:00h) e outro à tarde (13:00h às 16:30h). Por fim, no perfil 3, foi identificado um período de pico da manhã bem acentuado entre 06:45h e 07:50h, semelhante ao pico identificado no perfil 2. O mesmo ocorre no período de entropico, onde foram identificados dois agrupamentos, diferindo em menos de uma hora entre os dois perfis. O período de pico da tarde/noite do perfil 3 também se assemelha bastante aos perfis 1 e 2, diferindo em 20 minutos.

De posse dos horários de pico definidos, foram geradas as séries de volume dentro dos horários para todos os equipamentos. Na figura 3 é apresentado o resultado da série temporal para o horário de pico da manhã de um dado equipamento, classificado dentro do perfil 3, juntamente com os *outliers identificados* e o histograma da variável. O teste de normalidade de Shapiro-Wilk, com $\text{valor-}p = 0,52$, mostrou haver evidências de que a distribuição da variável volume de tráfego não difere significativamente da distribuição normal. O teste foi realizado para todos os equipamentos dentro desse perfil e a hipótese nula de normalidade não pôde ser rejeitada na maioria dos testes, ao nível de significância de 5%. Quanto à série temporal, nota-se que ela não apresenta grandes variações no volume na maior parte do ano.

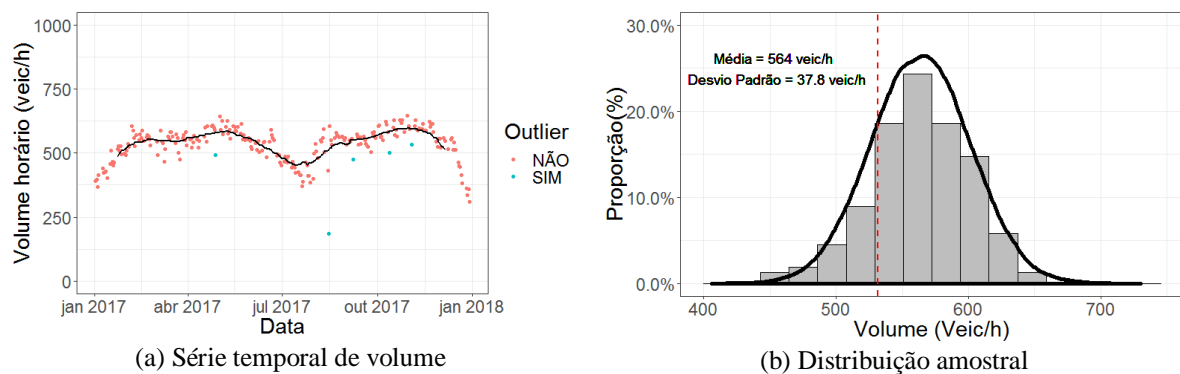
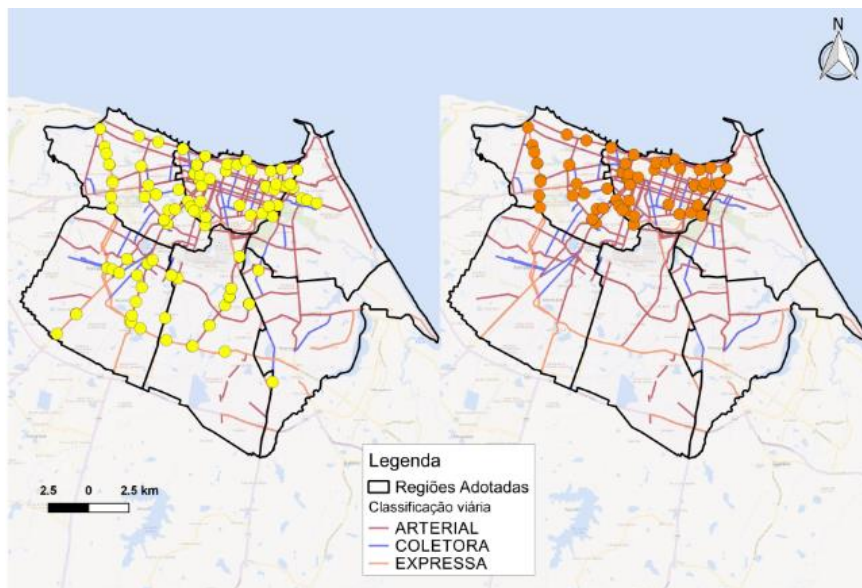


Figura 3: Série e distribuição amostral do volume veicular para o pico manhã

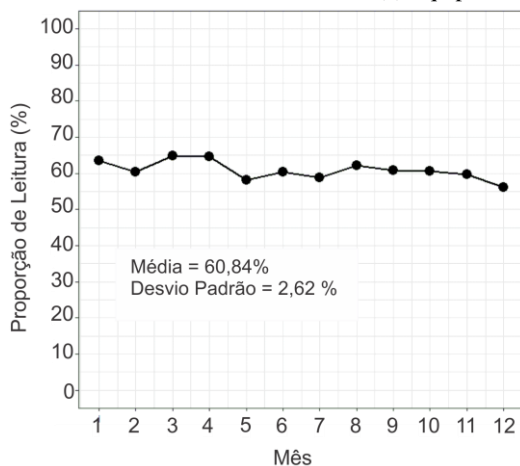
4.2. Extração do Fluxo OD

A etapa de extração de fluxos OD se iniciou pela seleção de equipamentos em função da variação da proporção de leituras ao longo do período de interesse. Após realização da análise de sensibilidade, foi definido como critério máximo para aceitação de um equipamento, o valor de 13% para o desvio padrão de proporção de leituras, resultando em um total de 106 equipamentos para a realização da associação entre placas, indicados na Figura 4a, distribuídos no zoneamento definido para a área de estudo, com destaque para as regiões Noroeste e Central. Estas regiões foram escolhidas para ilustrar os resultados desta etapa por conta da maior concentração de equipamentos na sua área. Já as Figuras 4b e 4c apresentam a variação da proporção de leituras ao longo do ano de 2017 para um equipamento que foi selecionado e outro que foi descartado, respectivamente.

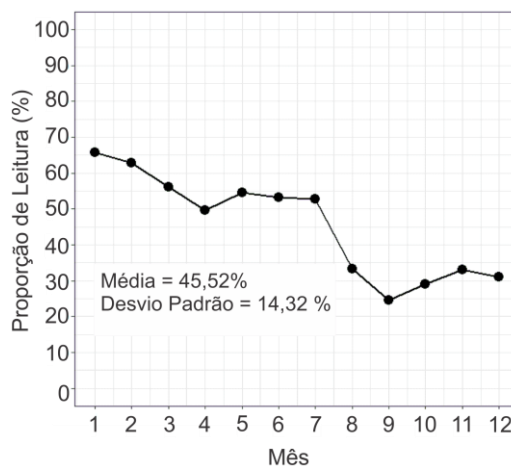
A partir dos resultados da associação, a última etapa anterior à geração das séries consistiu no filtro dos deslocamentos em função do tempo de viagem observado. Por fim, dado que os deslocamentos restantes entre pares OD identificados são considerados representativos da realidade no que diz respeito à sua variação ao longo do período de estudo, foram geradas as séries temporais de fluxo OD dentro dos horários para todos os pares de regiões e analisada a distribuição da variável. Na figura 5 é apresentado o resultado da série no horário de pico da manhã para as viagens entre as regiões noroeste e central, juntamente com os *outliers identificados* e o histograma da variável. O teste de normalidade de Shapiro-Wilk, com $valor-p = 0,19$, mostrou haver evidências de que a distribuição da variável fluxo OD não difere significativamente da distribuição normal. O teste foi realizado para todos os pares OD da área de estudo, e a hipótese nula de normalidade não pôde ser rejeitada na maioria dos testes, ao nível de significância de 5%. Quanto à série temporal, nota-se claramente a presença de uma sazonalidade das viagens entre essas duas regiões no pico da manhã ao longo do ano, sendo observada uma queda na quantidade de viagens nos meses de janeiro, julho e dezembro.



(a) Equipamentos selecionados na área de estudo



(b) Equipamento selecionado



(c) Equipamento não selecionado

Figura 4: Equipamentos selecionados e proporção de leituras no ano de 2017

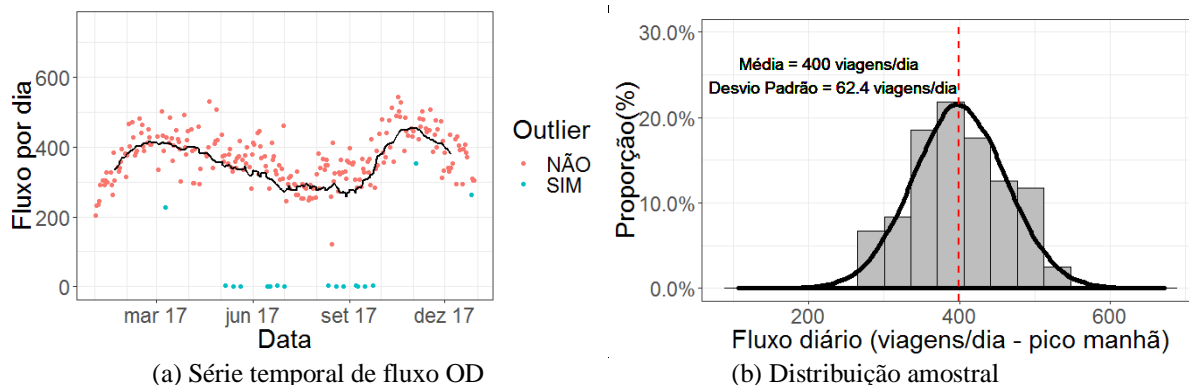


Figura 5: Distribuição e série do fluxo OD veicular obtido no horário de pico da manhã

5. CONSIDERAÇÕES E RECOMENDAÇÕES

Este trabalho apresentou uma metodologia de consolidação de dados de FE para a extração de séries temporais das variáveis volume de tráfego e fluxo OD. O estudo contribuiu para o tratamento de dados para geração de séries temporais que permitem avaliar a variabilidade do tráfego no meio urbano, subsidiando assim estudos que visem verificar de forma empírica premissas sobre modelos teóricos de estimação de matrizes OD e de alocação dinâmica dia-a-dia do tráfego veicular.

Nesse sentido, é essencial antes de realizar qualquer análise de dados provenientes de sistema de coleta automática de tráfego, a aplicação de um tratamento dos dados para identificar dados suspeitos (falhas de equipamentos), dados anormais (devido a eventos atípicos como acidentes, obras e outros), definir períodos representativos do tráfego (pico, entrepico, e fora pico), e definir viagens representativas entre zonas do meio urbano. Embora alguns dos critérios de seleção de equipamentos tenha sido em parte arbitrários, o trabalho contribuiu principalmente na definição e análise dos períodos de pico do tráfego, aplicando técnicas de agrupamento, que para o caso de Fortaleza-CE, permitiram avaliar a tendência de viagens pendulares em direção à região central da cidade, onde se concentra a maior parte dos empregos e atividades comerciais e de serviços.

O estudo também contribuiu para utilização de dados de sistema de Fiscalização Eletrônica para geração de fluxos de viagens entre zonas, que permitem analisar a variabilidade da variável fluxo OD que é essencial para o planejamento urbano de grandes cidades, mas que ainda é de difícil obtenção devido ao grande esforço de coleta. Ressalta-se, portanto, o esforço neste estudo de definir regiões de agregação dos dados (definindo um nível de agregação que fosse adequado para gerar viagens entre regiões conforme o nível abrangência da distribuição espacial dos equipamentos de fiscalização), de selecionar equipamentos de detecção e registro de placas com baixa variação ao longo do ano na proporção de leituras, e, além disso, o esforço de definir viagens entre regiões com base em tempos de viagem que representassem os deslocamentos entre as regiões de estudo.

Como limitações, identificou-se que vários equipamentos não estariam aptos para geração de séries de volume e fluxo OD, evidenciando, portanto, que ainda existem falhas na operação (e.g., falta de sincronia do relógio dos equipamentos, falha de leituras de placas e falha de registro da passagem dos veículos) que precisam ser ajustadas junto às empresas. Outra limitação do estudo diz respeito à distribuição de equipamentos na área analisada, onde existem poucos equipamentos em algumas regiões, localizados em vias arteriais da rede.

Agradecimentos

Os autores agradecem o apoio financeiro da CAPES e do CNPq para o desenvolvimento deste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

- Bertini, R. L.; M. Lasky e C. M. Monsere (2005) Validating predicted rural corridor travel times from an automated license plate recognition system: Oregon's frontier project. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, v. 2005, p. 706–711.
- Cascetta, E. (2009) *Transportation System Analysis: Models and applications*. Ed. Springer (2ª ed.), New York, USA.
- Castillo, E.; J. M. Mennéndez e P. Jimenez (2008) Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. *Transportation Research Part B: Methodological*, v. 42, n. 5, p. 455–481.
- Cheng, T.; J. Haworth e J. Wang (2012) Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems*, v. 14, n. 4, p. 389–413.
- Chu, K. K. A. e R. Chapleau (2010) Augmenting Transit Trip Characterization and Travel Behavior Comprehension. *Transportation Research Record: Journal of the Transportation Research Board*, v. 2183, p. 29–40.
- Cremer, M. e H. Keller (1987) A New Class of Dynamic Methods for the Identification of Origin-Destination Flows. *Transportation Research B*, v. 21, n. 2, p. 117–132.
- Ester, M.; H. P. Kriegel; J. Sander e X. Xu (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD-96 Proceedings*, p. 226–231.
- Giraud, T. (2018) OSRM: Interface Between R and the OpenStreetMap-Based Routing Service OSRM. R package version 3.1.1. <https://CRAN.R-project.org/package=osrm>.
- Oliveira, M. V. T. e C. F. G. Loureiro (2006) Análise dos Padrões de Variação Espaço-Temporal do Volume Veicular no Ambiente Urbano de Fortaleza. *Anais do XX Congresso de Pesquisa e Ensino em Transportes, ANPET, Brasília*, v. 1, p. 149–161.
- Pitombeira-Neto, A. R. (2015) Dynamic bayesian statistical models for the estimation of the origin-destination matrix. Tese de Doutorado – Programa de Pós-Graduação em Engenharia de Transportes, Departamento de Engenharia de Transportes, Universidade Federal do Ceará, Fortaleza, Brasil.
- Pitombeira-Neto, A. R. e C. F. G. Loureiro (2016) A Dynamic Linear Model for the Estimation of Time-Varying Origin–Destination Matrices from Link Counts. *Journal of Advanced Transportation*, v. 50, n. 8, p. 2116–2129.
- Pitombeira-Neto, A. R.; C. F. G. Loureiro, e L. E. Carvalho (2018). Bayesian Inference on Dynamic Linear Models of Day-to-Day Origin-Destination Flows in Transportation Networks. *Urban Science*. v. 2, n. 4, p. 117.
- Roess, R. P. e W. R. McShane (2004) *Traffic Engineering*, Ed. Pearson/Prentice-Hall, New Jersey, USA.
- Stathopoulos A. e M. G. Karlaftis (2001) Temporal and Spatial Variations of Real-Time Traffic Data in Urban Areas. *Transportation Research Record: Journal of the Transportation Research Board*, Washington, D.C., USA, n. 1768, p. 135–140.
- Weijermars, W. A. M. (2007) Analysis of urban traffic patterns using clustering. Netherlands TRAIL Research School.
- Zhao, J.; A. Rahbee e N. H. M. Wilson (2007). Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. *Computer-Aided Civil and Infrastructure Engineering*, v. 22, n. 5, p. 376–387.

Joana Maia Fernandes Barroso (joana@det.ufc.br)
João Lucas Albuquerque Oliveira (joao@det.ufc.br)
Francisco Moraes de Oliveira Neto (moraes@det.ufc.br)
Departamento de Engenharia de Transportes, Centro de Tecnologia, Universidade Federal do Ceará.
Campus do Pici, s/n – Bloco 703 Fortaleza, CE, Brasil.