

Avaliação das mudanças nas velocidades das linhas de ônibus da cidade de São Paulo após a implantação de faixas exclusivas através da análise de dados de GPS

Renato Oliveira Arbex¹ e Claudio Barbieri da Cunha²

Resumo: Este artigo apresenta uma análise da evolução das velocidades observadas nas linhas de ônibus do município de São Paulo entre os períodos de agosto de 2013 e abril de 2015. Nesse período foram implantadas faixas exclusivas para ônibus pela cidade que buscam priorizar o transporte coletivo nessas vias. Para tanto, foi desenvolvido um algoritmo que permitiu calcular as velocidades das viagens dos ônibus nas linhas através dos dados de GPS dos veículos. O algoritmo foi aplicado a um conjunto de linhas de ônibus do transporte coletivo municipal para um período de 20 dias úteis no mês de agosto de 2013, ou seja, antes da implantação das faixas exclusivas, e também para abril de 2015, quando as mesmas já estavam em operação. Dois grandes grupos de linhas foram analisados: 68 linhas que não tiveram faixas exclusivas implantadas em seus itinerários (não beneficiadas) e 240 das que tiveram ao menos um trecho dos seus respectivos trajetos com faixa exclusiva implantada. Os resultados obtidos mostram que, enquanto as linhas não beneficiadas tiveram um aumento médio de apenas 0,39% na mediana das velocidades das viagens com partidas entre 18h-19h, ou seja, no período mais crítico de pico da tarde, as linhas do grupo beneficiado tiveram aumento expressivo médio de 9,13% na mediana das velocidades. Entretanto, foi também observado que a distribuição das melhorias nos tempos de viagem é heterogênea, ou seja, algumas linhas tiveram redução das velocidades dos seus trajetos neste período, apesar dos benefícios das faixas exclusivas.

Palavras-chave: faixas exclusivas de ônibus; dados de GPS; velocidades de ônibus; monitoramento.

Abstract: This paper presents an analysis of the evolution of the observed speeds in bus lines of the bus service of the city of São Paulo between the periods of August 2013 and April 2015. In that period, a series of dedicated bus lanes were implemented across the city with the aim to prioritize the public transportation in those roads. To accomplish this, we have developed an algorithm that enables us to calculate travel times of bus lines through data from on board GPS devices. The algorithm was applied to 20 weekdays in August 2013, prior to the implementation of the dedicated bus lanes, and again in April 2015, when the set of these lanes was fully operational. Two large sets of bus lines were selected to be analyzed: 68 lines that did not have any relation to dedicated lanes, and 240 bus lines that had at least a small length of their respective itineraries along dedicated bus lanes. The results show that, while the non-benefited lines had a slight average increase of 0,39% in median speeds for trips departing between 18h-19h period, i.e., in the most critical evening peak, lines from the benefited group had an expressive increase of 9,13% in their speeds. However, we also observed that the distribution of speed improvements is uneven, and some lines had reduction of median speeds during this period, in spite of dedicated bus lane benefits.

Keywords: exclusive bus lanes; avl data; bus speeds; monitoring.

1. INTRODUÇÃO

A cidade de São Paulo atualmente conta com cerca de 11,9 milhões de habitantes (IBGE, 2016) e uma frota de mais de 8,1 milhões de veículos (DETRAN-SP, 2016). Em 2013, 120 mil novos veículos se juntaram à atual frota, o maior acréscimo anual até então observado (DETRAN-SP, 2016). Entretanto, a malha viária da cidade tem se mantido praticamente a mesma, com escassas intervenções para ampliação de capacidade ou construção de novas vias. Adicionalmente, o novo Plano de Mobilidade de São Paulo visa reduzir o uso do transporte individual, e devido a este novo foco a construção de novas vias é desestimulada (SMT, 2016), tendo em vista os investimentos elevados requeridos e os benefícios socioeconômicos limitados e praticamente nulos quando medidos em um horizonte de tempo maior;

em outras palavras, novas vias tendem a congestionar rapidamente como consequência da atratividade para o automóvel particular. Sendo assim, o crescimento do congestionamento no centro urbano é inevitável caso não sejam tomadas mais medidas para reduzir o uso do automóvel.

No dia 23 de maio de 2014, a cidade de São Paulo registrou o maior índice de congestionamento da sua história no período vespertino, às 19h de uma sexta-feira, com um total de 344 km dos 868 km monitorados pela Companhia de Engenharia de Tráfego (Folha de São Paulo, 2014). Esses congestionamentos já se tornaram rotina na cidade, pois a média diária no horário referido é de 218 km (Folha de São Paulo, 2014). São Paulo é a sétima metrópole mais congestionada do mundo de acordo com um índice de congestionamento elaborado pela empresa de monitoramento de tráfego em tempo real Tomtom (2014) que mede o acréscimo dos tempos médios de viagem nos horários de pico em relação aos tempos nas horas fora dos picos. Como consequências dos congestionamentos, pode-se citar uma menor qualidade de vida da população devido ao aumento do estresse (Hennessy e Wiesenthal, 1997), maior emissão de poluentes de automóveis e ônibus devido às baixas velocidades (André e Hammarström, 2000) e impactos na atividade econômica devido ao aumento dos custos de transporte (Weisbrod *et al.*, 2001).

¹ Renato Oliveira Arbex, Departamento de Engenharia de Transportes, USP. (renatoarbex@usp.br)

² Claudio Barbieri da Cunha, Departamento de Engenharia de Transportes, USP. (cabcunha@usp.br)

Manuscrito recebido em 21/09/2015 e aprovado para publicação em 26/05/2016.

Este artigo é parte de TRANSPORTES v. 24, n. 4, 2016. ISSN: 2237-1346 (online). DOI:10.14295/transportes.v24i4.1008

Os congestionamentos contribuem para a redução da qualidade de vida da população ao concentrar a poluição do ar nas regiões de maior congestionamento, trazendo problemas respiratórios. Além disso, o estresse é um dos grandes problemas causados aos motoristas, tanto pela poluição sonora como pelo grande tempo gasto no trânsito em seus deslocamentos. Frente a tais problemas a Prefeitura de São Paulo tomou medidas, nos últimos anos, para melhorar as condições do transporte público, mesmo que isto viesse afetar o transporte individual por automóvel.

O transporte público urbano é parte essencial de uma cidade. Idealmente deve constituir o meio de locomoção primário, garantindo o direito de ir e vir de seus cidadãos. Usando o sistema de transporte coletivo, a população contribui para a diminuição dos congestionamentos ao não ocuparem o espaço viário caso utilizassem automóvel, além de reduzir a poluição do ar e o consumo de combustíveis fósseis.

Sendo assim, o plano de melhoria do transporte público da Prefeitura de São Paulo visou a criação de faixas exclusivas de ônibus, atingindo em 2014 a marca de 369,7 km (CET-SP, 2015). A questão que surge é: será que essa implantação de faixas de ônibus contribuiu de fato para aumentar a velocidade das linhas e, conseqüentemente, reduzir os tempos de viagem dos usuários do transporte coletivo?

Este artigo objetivo relatar os resultados da pesquisa em que dados de monitoramento GPS de toda a frota do transporte público de São Paulo são analisados, antes e depois da instalação das faixas exclusivas de ônibus, para verificação das alterações das velocidades das viagens das linhas de ônibus da cidade. Em São Paulo, são ao todo mais de 1300 linhas operando com mais de 14 mil veículos na cidade (SPTTrans, 2015a), que emitem informação de localização por GPS com uma certa periodicidade (a cada 85 segundos em 2013 e a cada 40 segundos em 2015), gerando 12 milhões de informações de GPS diárias em 2013 e 24 milhões de registros diários em 2015. Devido a esse grande volume de dados, foi desenvolvida e aplicada uma metodologia de análise específica e eficiente para tratá-los, que é descrita na Seção 3, imediatamente após a seção em que são revisados os principais trabalhos da literatura científica que tratam do tema. Em seguida são apresentados, na Seção 4, os resultados obtidos e, finalmente, as considerações finais.

2. USO DE DADOS DE GPS PARA CÁLCULO DE TEMPOS DE VIAGEM NO TRANSPORTE PÚBLICO

Chakroborty e Kikuchi (2004) investigaram a utilização de dados de GPS de veículos para calcular os tempos de viagem dos ônibus e utilizar essa informação para criar modelos que possam prever os tempos de viagem de automóveis em corredores urbanos. Os tempos de viagem de veículos e ônibus foram analisados simultaneamente para um corredor urbano em Delaware, EUA. Os autores relataram que a diferença de tempos de viagem foi estável e foram criados modelos para prever os tempos de viagem do corredor com base nos GPS dos ônibus.

Dados de GPS estão sendo cada vez mais utilizados para determinação de velocidades médias e tempos de viagem de serviços de ônibus. Cortés *et al.* (2011) aplicou uma

metodologia que permitiu calcular velocidades médias das linhas de ônibus da cidade de Santiago, no Chile. Os autores comentam sobre o desafio de processamento de grandes volumes de dados de uma forma sistemática. A metodologia foi aplicada a uma semana de dados de GPS nas 700 linhas que operam em Santiago, emitindo dados de GPS a cada 30 segundos por veículo. A metodologia desenvolvida pelos autores pode ser aplicada a cada linha, ou a segmentos de extensão arbitrária.

Trétreault e El-Geneidy (2010) utilizam dados armazenados de localização automática de veículos e de bilhetação eletrônica de uma linha de ônibus, em conjunto com dados desagregados de uma pesquisa de comportamento de viagem para estimar o tempo de viagem de um novo serviço de linha de ônibus expressa que atuará em conjunto com uma linha de alta demanda já existente. Os dados de GPS são usados em um modelo que considera quantidade de embarques para estimar o ganho de tempo de viagem ao implantarem a nova linha. Os resultados possibilitaram definir os pontos de parada desse novo serviço de modo a minimizar os tempos de viagem dos usuários.

Ao analisar os tempos de viagem, é de grande importância olhar também para a questão da sua variabilidade. O conceito de variabilidade nos tempos de viagem refere-se à diferença entre tempos de viagem de diferentes viagens pelos veículos das linhas. Mazloumi *et al.* (2010) investigam a variabilidade ao analisar um conjunto de dados de GPS para uma única linha na cidade de Melbourne, na Austrália. Os autores relatam que alguns fatores levam a uma maior variabilidade nos tempos de viagem, como o uso do solo, extensão da rota, número de semáforos no trajeto e número de pontos de parada. Ressaltam ainda que a pouca disponibilidade de dados sobre tempos de viagem de linhas de ônibus e a limitação das metodologias de análise de dados de GPS fazem com que esta área de pesquisa ainda não tenha sido muito explorada na literatura. Assim, com o avanço da disponibilidade de dados de rastreamento e desenvolvimento de novas metodologias, o estudo da variabilidade dos tempos de viagem parece ser uma área promissora de pesquisa, apontam os mesmos.

Mais recentemente, Ye *et al.* (2016) analisam a variabilidade dos tempos de viagem de viagens de ônibus urbanos ao longo de um corredor de transporte na cidade de Chengdu, China. Medidas baseadas nos percentis das distribuições dos tempos de viagem foram usadas para investigar a variabilidade. Os autores também fizeram modelos de regressão de modo a identificar os principais fatores que afetam a variabilidade dos tempos de viagem. Os fatores incluíam, dentre outros analisados, a existência ou não de faixas exclusivas. Os dados de GPS dos ônibus são analisados para calcular o tempo de viagem dos veículos nas linhas. Os resultados sugerem que a existência de faixas exclusivas é mais eficiente para reduzir a variabilidade dos tempos de viagem nos horários de pico que em outros períodos do dia.

3. METODOLOGIA E APLICAÇÃO

Os dados de entrada para o algoritmo de análise são os dados de GPS da frota de ônibus de São Paulo, assim como os dados referentes à rede de transporte público, incluindo as informações de localização de pontos de ônibus, frequências e trajetos das linhas. Os mesmos foram obtidos

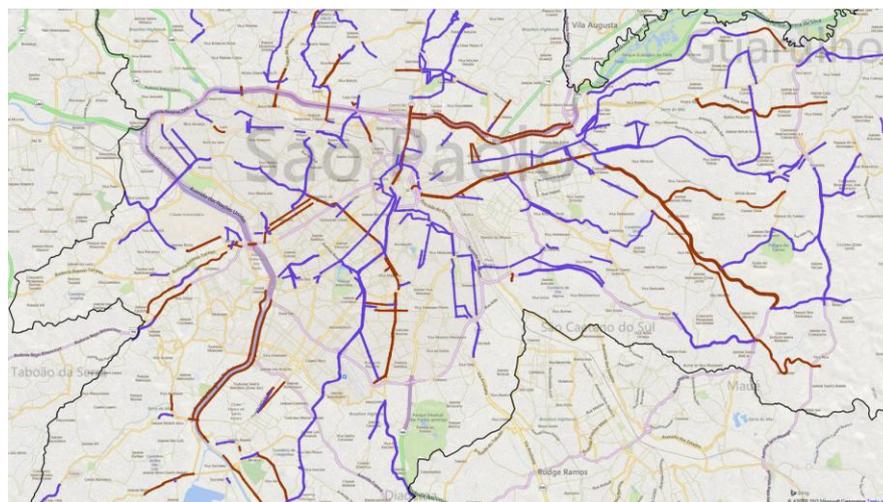


Figura 1. Mapa das faixas exclusivas de ônibus em São Paulo

junto à empresa responsável pelo planejamento e fiscalização da operação das linhas de ônibus urbanas de São Paulo, a SPTrans. A empresa disponibiliza uma página da internet específica para desenvolvedores terem acesso aos dados referentes à rede de transporte público mencionados e os dados de GPS dos veículos em tempo real (SPTrans, 2015b). Por terem um volume maior, os dados utilizados neste trabalho foram obtidos durante o evento “Hackatona do Ônibus” (Prefeitura de São Paulo, 2016), realizado em outubro de 2013 e diretamente com um disco rígido externo na SPTrans, para os dados referente ao mês de abril 2015.

A base de dados das linhas utilizado está no formato GTFS (*General Transit Feed System*), que é um formato padrão muito utilizado atualmente em diversas cidades do mundo para a distribuição das informações de seus sistemas de transporte ao *Google Transit* (Google Transit, 2015), assim como a disponibilização das informações ao público em geral e a desenvolvedores interessados em criar aplicações de transporte.

De 2013 ao final de 2014, foram implantados 369,7 km de faixas exclusivas de ônibus, sendo a maior parte entre os meses de agosto de 2013 e final de 2014. Estas faixas são preferenciais para o transporte coletivo por ônibus e operam à direita das vias. A Figura 1 mostra as faixas exclusivas para ônibus da cidade de São Paulo. Em roxo estão representadas as faixas implantadas durante esse período de análise, entre agosto de 2013 e abril de 2015, e aquelas em cor castanho avermelhado as que já haviam sido implantadas anteriormente.

Os dados que foram possíveis de serem obtidos e utilizados para análise foram dois meses de dados de GPS, o primeiro mês foi agosto de 2013 e o segundo mês foi abril de 2015. O ideal seria uma comparação entre os mesmos meses do ano, contudo, existe uma dificuldade de obtenção desses dados operacionais com uma frequência mensal diretamente com a empresa, de tal forma que dados para outros períodos não estavam disponíveis para a realização da pesquisa. De toda forma, os meses escolhidos representam um perfil semelhante de demanda, já que em agosto de 2013 foram transportados 261 milhões de passageiros em 22 dias úteis e em abril de 2015 foram 240 milhões em 20 dias úteis (PMSP, 2016), uma variação de 8%, que pode ser explicada tanto pela diferença de 2 dias úteis, pelo feriado de Tiradentes (21 de abril de 2015) ter ocorrido numa 3ª feira, afetando o movimento de passageiros no dia útil anterior, e também por abril ter um dia a menos que agosto.

Para a análise neste artigo, foram considerados os dias úteis destes meses, mais relevantes em termos tanto de ocorrência de congestionamentos e, portanto, efetividade das faixas exclusivas, quanto de número de passageiros beneficiados. Ao todo foram considerados 20 dias úteis em agosto de 2013 e igualmente 20 dias úteis em abril de 2015. Os dados de GPS disponibilizados contêm 6 informações. As informações para cada registro são:

- i. **dataserver**: data e hora exata que o servidor recebeu a informação do GPS;
- ii. **dataavl**: data e hora exata que o AVL (*automatic vehicle location*, o aparelho de GPS) enviou a informação ao servidor;
- iii. **linha**: código da linha conforme cadastro interno do sistema OlhoVivo da SPTrans;
- iv. **lat**: latitude da informação, com 6 casas decimais;
- v. **lon**: longitude da informação, com 6 casas decimais;
- vi. **codavl**: código do equipamento de AVL do veículo (representa um veículo único).

A seguir é exibido um exemplo de registro:

```
“2014-10-20 08:18:10.460,2014-10-20  
08:18:07.000,478,-23.589373,-4
```

3.1. Escolha das Linhas para os Grupos de Comparação: Beneficiadas e Não-Beneficiadas com as Faixas Exclusivas

Devido ao grande número de faixas exclusivas para ônibus que foram implantadas na cidade de São Paulo e a sua distribuição não homogênea no espaço viário, decidiu-se que a forma mais adequada de análise é através da comparação das medianas das velocidades das viagens das linhas de ônibus da cidade. Dois grupos de linhas foram analisados: um grupo com as linhas que tiveram ao menos um trecho de seu itinerário beneficiado com a implantação de faixa exclusiva e outro grupo com as linhas que não tiveram faixas exclusivas aplicadas em seus trajetos.

A seguir são detalhados os procedimentos para a escolha dos conjuntos de linhas a serem comparadas entre 2013 e 2015.

Operavam na cidade de São Paulo, em abril de 2015, 1523 linhas do transporte coletivo urbano municipal (PMSP, 2016). A fim de racionalizar o esforço de avaliação dos impactos das faixas exclusivas, foi inicialmente realizada uma análise da representatividade do conjunto de linhas em termos de participação na demanda. Os resultados

dessa análise indicam que do total de 1523 linhas, aproximadamente apenas 200 linhas já representam 40% de toda a demanda de transporte por ônibus daquele mês na cidade; 600 linhas, que correspondem a menos da metade do total de linhas do sistema, são responsáveis por transportar 80% da demanda total.

Adicionalmente, das 1523 linhas em operação em abril de 2015, 923 linhas (que transportam 80,1% da demanda) foram beneficiadas com ao menos um trecho de faixa exclusiva nos seus respectivos trajetos, enquanto as demais 600 (que transportam apenas 19,9% da demanda) não o foram. Isso indica que as novas faixas exclusivas implantadas procuraram priorizar os itinerários das linhas de ônibus responsáveis pela maior parcela da demanda do transporte coletivo no município.

Por outro lado, devido a problemas tais como a compatibilidade entre os dados de agosto de 2013 e abril de 2015, nem todas as linhas foram passíveis de análise dos dados de GPS de seus veículos. Do grupo de 923 linhas beneficiadas pelos corredores, 142 tiveram algum tipo de alteração nos pontos iniciais ou finais entre 2013 e 2015. Das 781 linhas restantes do grupo, algumas linhas circulares existentes foram desconsideradas devido à incompatibilidade dos dados do sistema GTFS (*General Transit Feed System*, formato da base de dados do cadastro das linhas) e dos códigos de GPS: nas linhas circulares, o GTFS apresenta como uma única linha, enquanto que o GPS tem duas linhas correspondentes cadastradas, o que impossibilitou as análises. Assim, retirando as 163 circulares, restaram 618 linhas. Dessas, as 330 linhas com as maiores demandas, representando 71,6% da quantidade total de passageiros transportados, foram escolhidas para o grupo das beneficiadas com faixas exclusivas. Foram escolhidas as 330 (54% das linhas) com maiores demandas pois foram as que registraram acima de 200 mil embarques no mês de abril de 2015, um critério adotado de forma a retirar a influência de linhas com baixa demanda na análise da evolução das velocidades, já que mesmo com grandes aumentos de velocidade essas linhas beneficiam proporcionalmente menos usuários (28,4%).

Nesse ponto foi possível observar que uma parte das linhas (82 das 330) tiveram mudanças nos itinerários ao longo dos dois anos e as distâncias entre os pontos iniciais e finais não eram as mesmas. Assim, apenas foram consideradas as linhas cuja distância do trajeto total não tivesse uma diferença maior que 1 km (medido conforme traçado representado no arquivo shapes.txt da especificação GTFS) entre 2013 e 2015. Por fim, inconsistências tais como erros sistemáticos (por exemplo, quando o sentido da linha fornecido pelo registro GPS era diferente do percurso efetivamente sendo realizado) nos dados de algum ano impossibilitam análises para outras 8 linhas, gerando ao final 240 linhas comparáveis nesses dois períodos. Esse processo demonstra a dificuldade de comparação de tempos de viagem e outros indicadores de desempenho e qualidade em sistemas de transporte público, já que a rede acaba sendo muito mutável ao longo de um horizonte de tempo maior, de vários meses, e pequenas alterações impossibilitam a construção de uma série histórica de dados consistentes. As 240 linhas finais selecionadas representam 54% da demanda das 923 iniciais, o que garante boa representatividade, já que representa uma amostra equivalente a pouco mais da metade dos passageiros beneficiados por linhas que passaram a operar com faixas exclusivas.

Das 600 linhas não beneficiadas com faixas exclusivas, eliminando-se as que tiveram itinerários modificados, alterações nos pontos finais e iniciais e as circulares, restaram 176. A eliminação nesse caso foi maior, pois em abril de 2015 estavam operando as linhas da rede noturna, que são 151 linhas circulares que não operavam em 2013, pois foram criadas posteriormente. Das 176 restantes, foram escolhidas 100 linhas (para representar igualmente 54%, como a seleção das linhas beneficiadas), das quais, retirando-se aquelas com diferença na distância dos trajetos superior a 1 km, resultaram 76 linhas. Inconsistências prejudicaram a análise dos tempos para 8 linhas, restando, ao final, 68 linhas para o grupo das linhas comparáveis entre 2013 e 2015. As 68 linhas representam 48% da demanda

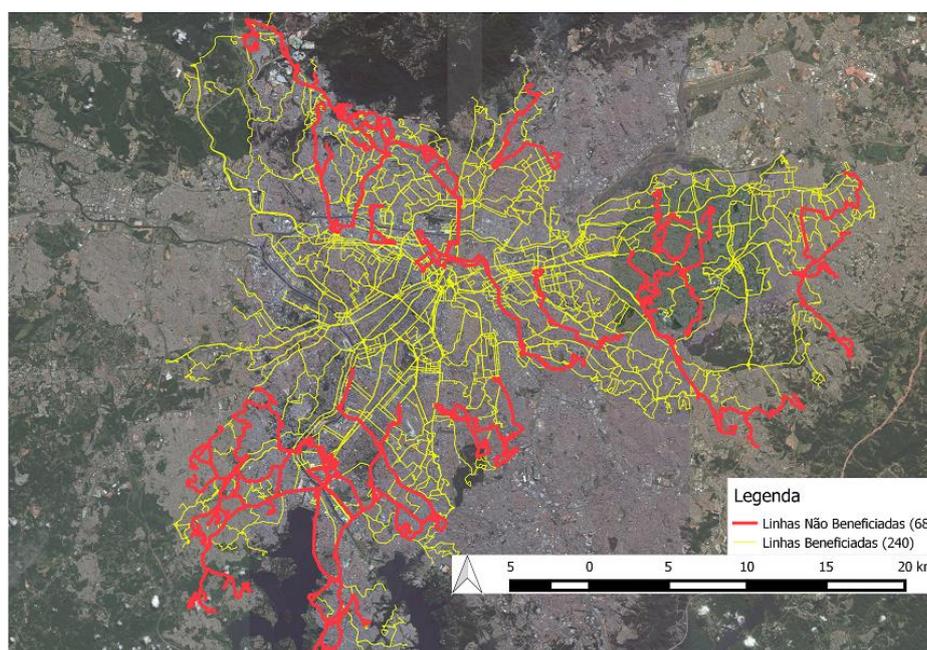


Figura 2. Linhas beneficiadas escolhidas em amarelo e não beneficiadas em vermelho

Tabela 1. Processo de escolha das linhas a serem consideradas para a comparação 2013-2015

Grupo	Total	Após Eliminação das alterações nos pontos iniciais ou finais	Retiradas as Circulares	Candidatas	Após retiradas as linhas com diferença na distância do trajeto (2013-2015) maior que 1km	Selecionadas (após retiradas linhas com inconsistências nos tempos de viagem)	% da demanda representada do total do grupo de linhas final escolhido
Beneficiadas	923	781	618	330	248	240	54%
Não-beneficiadas	600	271	176	100	74	68	48%

das 600 iniciais, aproximadamente a mesma representatividade do grupo das linhas beneficiadas pelas faixas exclusivas. As linhas selecionadas para a análise são mostradas no mapa da Figura 2, enquanto a Tabela 1 sumariza as linhas consideradas em cada fase do processo de seleção.

4. METODOLOGIA DE ANÁLISE DOS DADOS DE GPS PARA OBTENÇÃO DOS TEMPOS DE VIAGEM

Uma vez tendo sido selecionadas as linhas que compõem cada um dos dois grupos, as questões a serem investigadas, para cada grupo, são:

- Qual o padrão da mudança das velocidades das viagens entre agosto de 2013 e abril de 2015?
- O padrão de mudança no percentil 90 das velocidades (representando situações mais críticas) acompanhou o padrão de mudança das velocidades?
- Qual a economia de tempo entre esses dois períodos? Para tanto, foi desenvolvido um algoritmo de análise de dados de GPS que é detalhado a seguir nesta subseção.

Os dados referentes à rede de transporte utilizados são os arquivos do GTFS de São Paulo, que podem ser obtidos através da página na internet da SPTrans após um cadastro (SPTrans, 2015b). Os arquivos que foram usados nesta metodologia são:

- i. **trips.txt**: Descreve as viagens dos veículos, sendo uma em cada sentido da rota.
- ii. **stop_times.txt**: Contém a sequência dos pontos de parada de cada linha.
- iii. **frequencies.txt**: Arquivo com os intervalos programados para cada linha por horário.

- iv. **stops.txt**: Cadastro dos pontos de ônibus, terminais, estações de metrô e trem do sistema.
- v. **shapes.txt**: Localização exata do trajeto das linhas com a distância acumulada percorrida.

Para cada uma das linhas analisadas foi calculado o tempo de viagem total. De modo a evitar problemas de repetição de dados de GPS e veículos ligados nos terminais e pontos iniciais e finais das linhas, foi considerado o tempo de viagem entre o segundo e o penúltimo ponto de parada de cada linha, ou seja, excluindo-se os pontos terminais de extremidade das linhas. Os códigos destes pontos estão no arquivo stops.txt. Com a listagem das linhas de cada grupo e os códigos dos pontos iniciais e finais, foi aplicado o algoritmo que produz como resultado quaisquer percentis dos tempos de viagem entre os pontos escolhidos para todas as linhas. A partir da informação das quilometragens percorridas pelas linhas disponível no arquivo shapes.txt do GTFS foi possível determinar as velocidades observadas nesses percentis.

É importante salientar que alguns outros fatores exógenos, além da implantação das faixas exclusivas podem ter influenciado as mudanças de velocidades das linhas entre o período analisado, como eventuais obras nas vias, alterações no sistema viário ou algum tipo de interferência, sejam eventuais (como acidentes) ou frequentes (por exemplo, mudança na programação semafórica) nas vias durante algum dos meses analisados. Dessa forma, a presente análise procura avaliar a tendência de mudança no perfil das velocidades do conjunto do sistema de ônibus da cidade de São Paulo, sem levar em consideração esses fatores desconhecidos, para os quais não há qualquer tipo de registro que eventualmente permita considerá-los nesta análise.

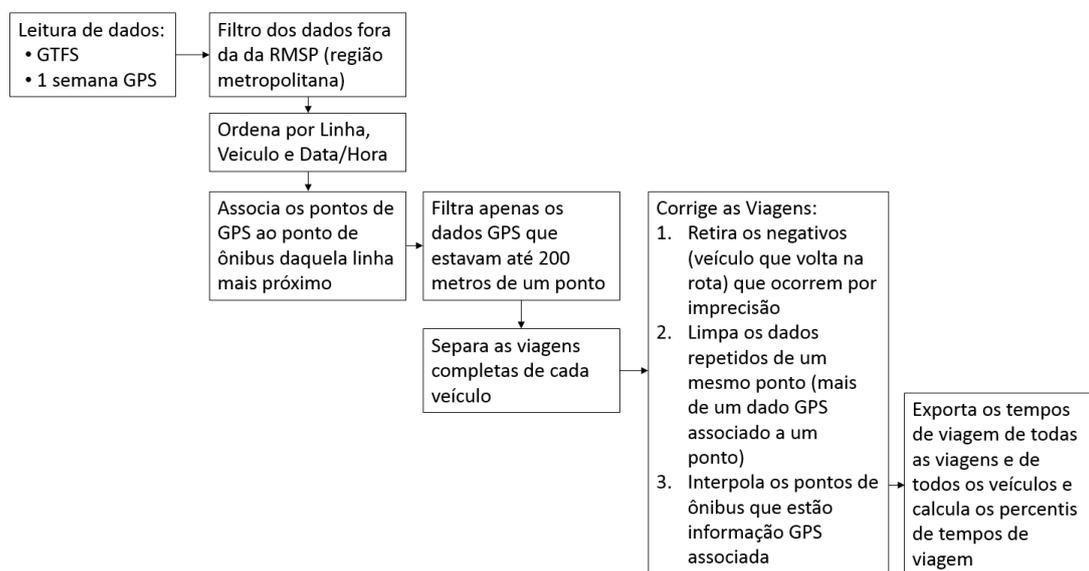


Figura 3. Fluxograma do algoritmo de processamento dos dados de GPS

O algoritmo proposto processa os dados de GPS da seguinte forma: resumidamente, cada registro de GPS é associado ao ponto de parada de ônibus da respectiva linha mais próximo espacialmente do mesmo. Dessa forma, determinam-se os horários de passagem de cada veículo de uma linha em cada um dos pontos de parada de ônibus ao longo da mesma. Finalmente, para os pontos de parada de uma linha que não tiveram nenhum dado de GPS associado, é feita uma interpolação para a continuidade dos dados de tempos de viagem entre todos pares de pontos de parada consecutivos. Seu funcionamento está resumido no fluxograma da Figura 3. Cada uma das suas etapas é detalhada a seguir.

- a. **Leitura de Dados:** Inicialmente são armazenados em estruturas de dados adequadas as informações dos pontos de ônibus (stops.txt), dos pontos por linha (stop_times.txt), dados de GPS, e as linhas a serem analisadas.
- b. **Filtro Espacial:** Na sequência, um filtro espacial é aplicado para evitar pontos de GPS com problemas de funcionamento onde ele emitiria uma localização errada, fora da região metropolitana da cidade de São Paulo (RMSP), por algum problema de erro de leitura de posicionamento do equipamento, levando à necessidade de descarte desses registros. Esse recorte espacial mais amplo que a cidade de São Paulo foi escolhido a fim de oferecer uma margem de segurança a mais no filtro, de forma a incluir todos os dados potencialmente adequados. Os dados de GPS, quanto a este quesito, são de boa qualidade, dado que, para uma semana de outubro de 2014, por exemplo, 99,989% dos dados são dentro da RMSP.
- c. **Ordenação:** Os dados de GPS são ordenados primeiramente por código de linha, em seguida por código de veículo e por último por data/hora. Desta forma, ordenados cronologicamente, os dados são preparados para uma etapa futura, de separação das viagens.
- d. **Associação ao Ponto de ônibus Mais Próximo:** nesta etapa, cada um dos dados de GPS é associado ao respectivo ponto de parada de ônibus mais próximo da respectiva linha do dado pela distância euclidiana. Desta forma, os pontos ficam com uma informação temporal da passagem de cada veículo de cada linha, sendo então possível o cálculo dos tempos de viagem entre os pontos.
- e. **Filtro Espacial de Associação ao Ponto Mais Próximo:** Os dados de localização dos veículos que estavam a mais de 200 metros do ponto mais próximo são descartados, já que a menor resolução espacial de interesse para a análise dos tempos de viagem é cada ponto de parada de ônibus da linha, pois os tempos de viagem poderão, dessa forma, serem obtidos para segmentos entre dois pontos consecutivos. Não podem ser considerados todos os pontos uma vez que registros de GPS que representem a localização de um veículo entre dois pontos de parada consecutivos não permite a associação do horário de passagem nos pontos de parada (se o fizermos, não representará o horário que ele efetivamente estava próximo daquele ponto). A quantidade de dados descartados depende da distância entre pontos. Por exemplo, na linha

8700-10, que percorre o corredor de ônibus da Avenida Rebouças, de 244.810 registros de GPS de uma semana de dados, 57,2% dos dados foram válidos para a próxima etapa.

- f. **Separação de Viagens:** Nesta etapa, viagens diferentes são separadas para serem analisadas distintamente. O critério de corte para uma viagem ser separada é quando ocorre uma queda expressiva na sequência do ponto de ônibus associado ao dado de GPS. O critério utilizado foi de 80% do total de pontos de parada da linha. Assim, por exemplo, caso uma sequência de registros estivesse associado ao ponto número 88 da linha, e em seguida o próximo registro de GPS representasse o ponto de parada número 2 da linha, ocorreria o corte para separação entre diferentes viagens, pois a diferença de 86 (88-2) representa queda de 97% no número sequencial do ponto de parada associado.
- g. **Correção das Viagens:** Alguns problemas após a etapa anterior de separação das viagens devem ser corrigidos. Na Figura 4 e na Figura 5 estão dois gráficos gerados de uma linha, para exemplificar visualmente as etapas de separação das viagens e correção dos problemas. No eixo das abscissas está a sequência de informações de GPS ordenadas cronologicamente. No eixo das ordenadas, está a sequência dos pontos de ônibus associados a cada informação. Assim, conforme o veículo emite dados, isso significa que ele está avançando no seu itinerário, pois a sequência vai aumentando. Por exemplo, numa sequência 2, 5, 6, 8, 10, 15, equivaleria a 6 dados de GPS onde o primeiro registro estava mais próximo do segundo ponto de ônibus da linha, a próxima informação associada ao quinto ponto, e assim sucessivamente.

Foram avaliadas as necessidades de correções e limpeza adicionais da base antes de utilizá-la. São eliminados os dados que representam veículos que aparentemente “voltaram”, ou retrocederam indevidamente no tempo; ou seja, um dado de GPS foi associado a um ponto de ônibus anterior a um que já foi associado; para os dados repetidos, apenas um (o que foi enviado mais próximo ao ponto dentre todos) é usado para associar ao ponto mais próximo. Por último, os pontos de ônibus que não tiveram uma informação associada têm um tempo de passagem associado por interpolação. Um exemplo visual das correções necessárias é mostrado na Figura 5.

- a) **Exporta Informações e Percentis:** São exportados os dados de tempos de viagem entre os pontos de interesse e para as linhas de interesse. Também são calculados e exportados os valores dos percentis 50 e 90 dos tempos de viagem entre os pontos e linhas de ônibus escolhidos. Esses percentis específicos foram escolhidos pois são os percentis usados em cálculos de índices de variabilidade dos tempos de viagem (FHWA, 2016).

O algoritmo foi implementado na linguagem *Python* (Python, 2016) em um computador com sistema operacional *Windows 8*, Processador AMD FX 8150, 3.6 GHz, 64 bits, com 16 Gb de memória RAM, utilizando as bibliotecas de análise e tratamento de dados *Pandas* (Pandas, 2016) e *Numpy* (Numpy, 2016), que se mostraram bem eficientes

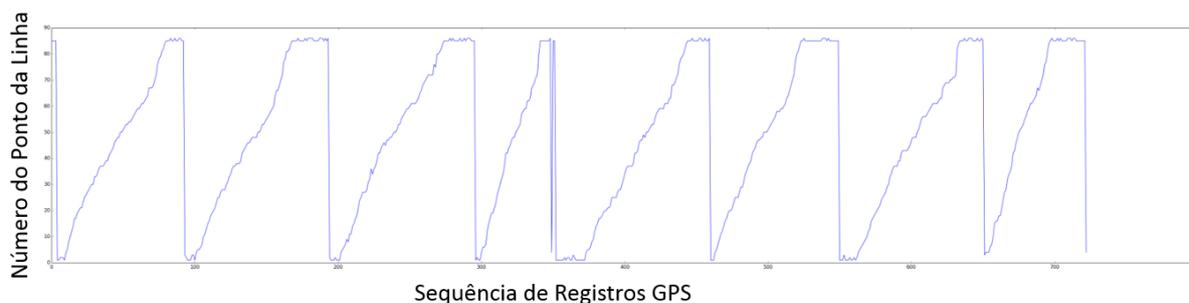


Figura 4. Exemplo de Sequência dos Pontos após a Associação ao Ponto mais próximo para um veículo

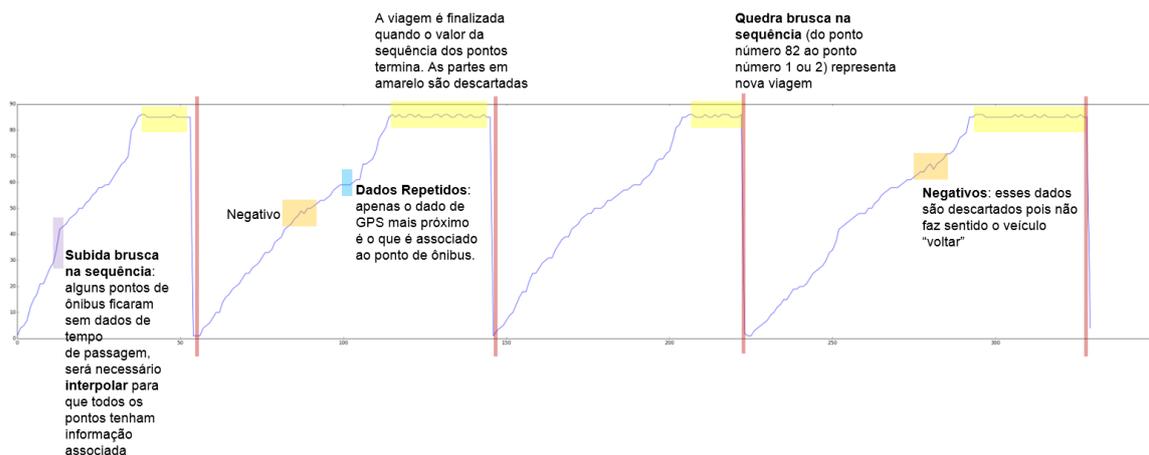


Figura 5. Exemplo de correções necessárias na sequência de dados de GPS de um veículo

para analisar o grande volume de dados. Ao total foram processados 60 Gigabytes de dados com 720 milhões de registros em aproximadamente 3 horas. Numa primeira etapa os arquivos que eram agrupados por dia foram separados em pouco mais de 2.600 arquivos representando cada linha. Para cada linha, foi criado esse arquivo tipo *csv* que tivesse todos os seus registros de dados de GPS. O arquivo tipo *csv* (“*comma separated values*”) representa um formato de arquivo com valores separados por vírgula. Desta forma, foi possível trabalhar com arquivos menores, evitando o problema de não conseguir ler todos os 20 dias úteis na memória do computador.

Como resultados do processamento dos dados de GPS pelo algoritmo, é gerada uma tabela principal de tempos de viagem em segundos, para qualquer combinação de dois pontos de parada de qualquer código de linha. Agregando e analisando esses dados podem ser gerados os percentis dos tempos de viagem para quaisquer pares de pontos selecionados de qualquer linha selecionada. Com os percentis dos tempos de viagem, é possível calcular indicadores de variabilidade e confiabilidade do tempo de viagem.

Foram calculados todos os tempos de viagem, para todas as viagens das linhas analisadas, ocorridas em todos os dias úteis de agosto de 2013 e abril de 2015. Ao todo foram 652 mil viagens registradas pelo algoritmo em agosto de 2013 e 741 mil em 2015. Com essas viagens foram calculados os percentis 50 e 90 por faixa horária, das partidas iniciadas entre 0h e 1h, 1h até 2h, assim sucessivamente até 23h-24h. Podem ser calculados quaisquer percentis de interesse, tendo sido escolhidos os percentis 50 e 90; o primeiro (50) a fim de permitir efetuar as comparações das medianas das velocidades médias das viagens e o percentil 90 por ser um indicador de tempos de viagens de situações mais críticas. O percentil 90 permite que avaliar se entre 2013 e 2015,

um eventual aumento das velocidades também foi acompanhado de uma redução dos tempos de viagem nos casos extremos das viagens com menores velocidades.

O período do dia escolhido para análise dos tempos de viagem foi das partidas que ocorreram entre 18h e 19h. Esse período de tempo foi escolhido tendo em vista não só os resultados obtidos por Ye *et al.* (2016), como também devido ao fato de que algumas faixas exclusivas em São Paulo só operam durante os horários de pico. Adicionalmente, nossos resultados indicam que esse é o período do dia onde as velocidades dos ônibus são as mais baixas, tendo em vista o pico da tarde apresentar mais congestionamentos na cidade.

Nesse período entre 18h e 19h foram identificadas 34.172 viagens nas 308 linhas em 2013 e 38.541 viagens em 2015. Acredita-se que o aumento do número de viagens detectado pelo algoritmo se deve ao fato da melhora da qualidade dos equipamentos GPS e da maior taxa de atualização, que passou ser de a cada 85 segundos em 2013, para 45 segundos em 2015.

5. RESULTADOS

A Figura 6 apresenta os gráficos *box-plot* das velocidades medianas verificadas para as 68 linhas sem benefícios de faixas exclusivas consideradas para análise. É possível verificar que não houve variação estatisticamente significativa da média para o grupo das linhas não beneficiadas, pois a média das velocidades medianas das viagens em 2015 continua no mesmo intervalo de 95% de confiança da média registrada em 2013 (12,85 km/h em 2013 e 12,9 km/h em 2015, representando 0,39% de aumento).

Esta evolução é diferente quando são analisadas as 240 linhas que foram beneficiadas com faixas exclusivas entre 2013 e 2015. Na Figura 7 está apresentado o gráfico *box-plot* das velocidades medianas dessas linhas. Neste caso, de fato a velocidade mediana aumentou de forma estatisticamente significativa, passando de 13,04 km/h em 2013 para 14,23 km/h para este grupo de linhas (9,13% de aumento).

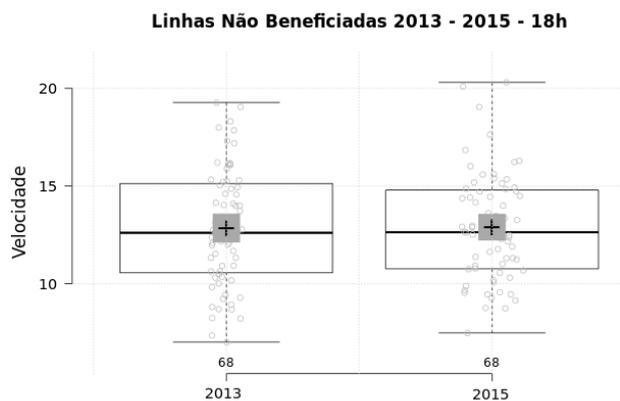


Figura 6. Evolução das Velocidades Medianas (percentil 50) das 68 linhas Não Beneficiadas por faixas exclusivas, entre 2013 e 2015, às 18h

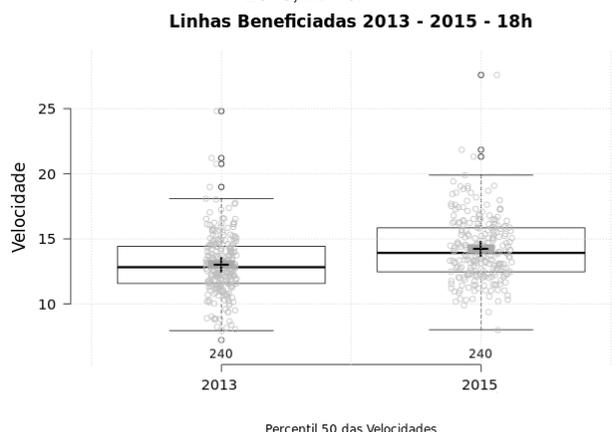


Figura 7. Evolução das Velocidades Medianas (percentil 50) das 240 linhas Beneficiadas por faixas exclusivas, entre 2013 e 2015, às 18h

Nas Figuras 8 e 9 estão apresentados os resultados que mostram a evolução das velocidades das partidas entre 18h e 19h para todas as linhas entre os períodos de agosto de 2013 e abril de 2015. Na Figura 8, estão as 68 linhas não beneficiadas. Em ambas as figuras, o eixo das abscissas representa a variação percentual da velocidade mediana (percentil 50 das velocidades das partidas), o eixo das ordenadas representa a velocidade mediana em 2015 para as viagens das linhas de ônibus cujas partidas ocorreram entre 18h-19h, e o tamanho da bolha representa o aumento ou diminuição da velocidade para o percentil 90. Ao representar a redução/aumento das velocidades no percentil 90 juntamente com a mudança no percentil 50, é possível visualizar que para algumas linhas ocorre uma melhora na velocidade mediana das viagens, mas uma piora nos casos extremos.

É possível observar que não se identificou um padrão definido. Uma parte das linhas teve um aumento de velocidade e outra parte teve uma redução. No geral, linhas que tiveram redução das velocidades correspondentes ao tempo

de viagem do percentil 50 também tiveram redução nas velocidades correspondentes ao percentil 90, ou seja, houve piora no dia a dia e piorou ainda mais nos dias críticos (bolhas vermelhas no lado esquerdo). Já as linhas com aumento da mediana das velocidades das viagens tiveram melhora nos casos extremos.

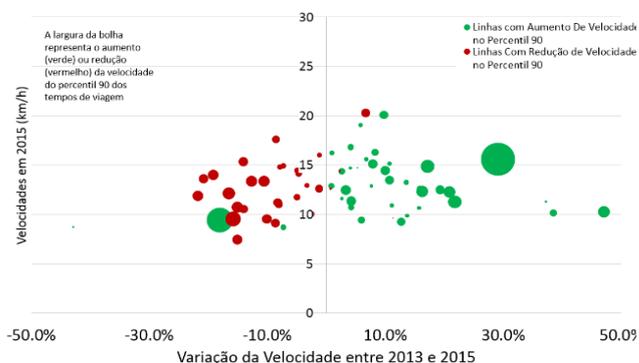


Figura 8. Evolução das Velocidades Medianas (percentil 50) e da Confiabilidade (pelo percentil 90) das 68 linhas Não Beneficiadas por faixas exclusivas, entre 2013 e 2015

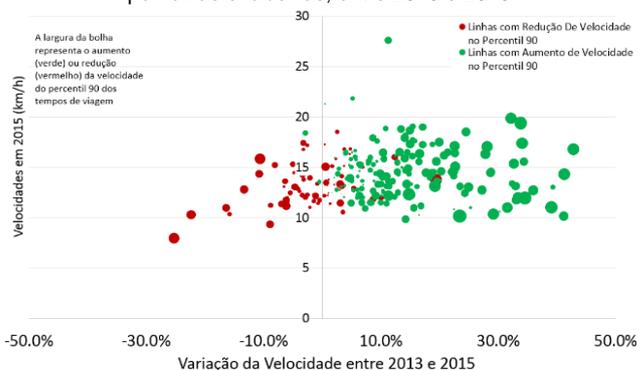


Figura 9. Evolução das Velocidades Medianas (percentil 50) e da Confiabilidade (pelo percentil 90) das 240 linhas Beneficiadas por faixas exclusivas, entre 2013 e 2015

Deve-se recordar que a Figura 9 representa as 240 linhas com maior demanda dentre as linhas beneficiadas que foram analisadas. É possível observar que há uma maior concentração de linhas na área à direita do gráfico, que representa ganho de velocidade mediana entre 2013 e 2015. Grande parte das linhas que tiveram melhora nas velocidades medianas também tiveram aumento nas velocidades no percentil 90 (que representam os 10% de partidas com menor velocidade no mês), representadas por bolhas verdes no lado direito do gráfico. Entretanto, existem linhas que, apesar de terem melhorado a velocidade correspondente ao percentil 50 dos tempos, tiveram redução de velocidade no correspondente ao percentil 90 (bolhas vermelhas do lado direito).

É interessante notar pelo gráfico da Figura 9 que praticamente todas as linhas operam com velocidades entre 10 e 18 km/h nas partidas realizadas entre 18h-19h nos 20 dias úteis em 2015. Apenas 3 das 240 linhas tem velocidades acima de 20km/h, e todas as 3 percorrem o trecho urbano da Rodovia Anhanguera em grande parte de seu trajeto.

Quanto aos benefícios das linhas com ao menos um trecho de faixa exclusiva, 29% das linhas deste grupo beneficiado tiveram redução de 15 minutos ou mais nos tempos de viagem de seus trajetos, enquanto que esse percentual foi de 8% para o grupo não beneficiado.

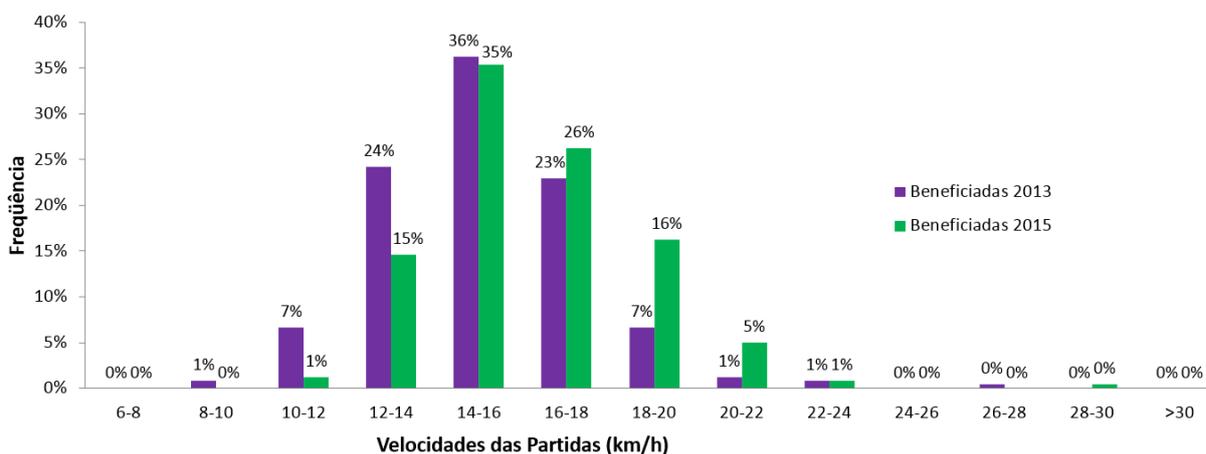


Figura 10. Velocidades Medianas (percentil 50) entre 2013 e 2015 para as Linhas Beneficiadas por Faixas Exclusivas das partidas às 18h-19h

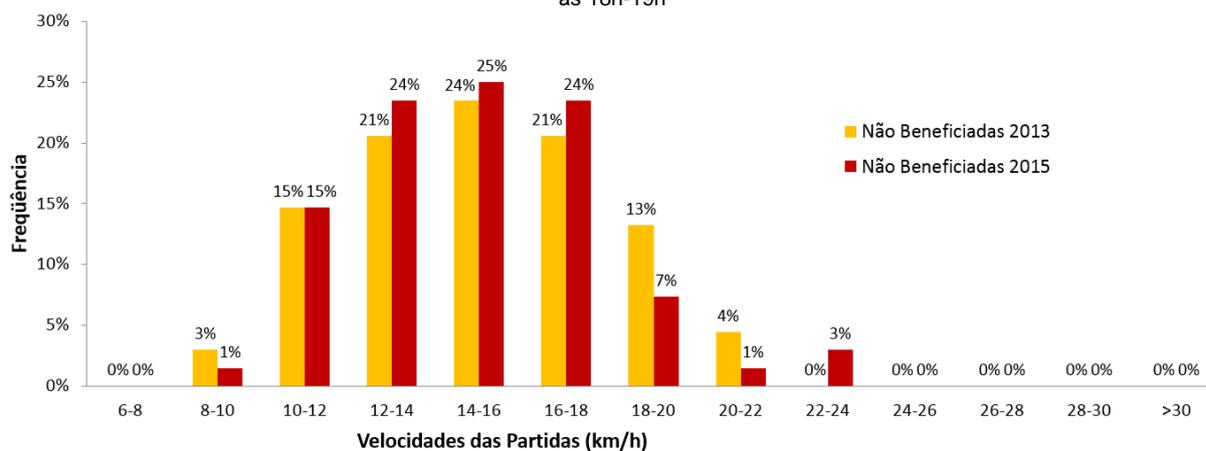


Figura 11. Velocidades Medianas (percentil 50) entre 2013 e 2015 para as Linhas Não Beneficiadas por Faixas Exclusivas.

Os valores médios de aumento de velocidade correspondente ao tempo de viagem para o percentil 50 foi de 9,13% para as linhas beneficiadas e 0,39% para as não beneficiadas. A economia de tempo de viagem de todo o percurso foi, em média, de 6 minutos e 46 segundos para as linhas beneficiadas e ganho médio de 1 minuto e 21 segundos para as não-beneficiadas.

As Figuras 10 e 11 apresentam, respectivamente, os histogramas correspondentes aos dois grupos de linhas, beneficiadas e não beneficiadas por faixas exclusivas, entre 2013 e 2015. De fato, houve um aumento maior do percentual de linhas com velocidades mais altas. Por exemplo, do grupo das não beneficiadas, em 2013, apenas para 38,2% das linhas a médias das velocidades medianas era maior que 16 km/h, valor que reduz para 35,3% em 2015. Já no grupo das linhas beneficiadas por faixas exclusivas, esse mesmo número passou de 32,0% para 48,8% das linhas, que representa uma maior porcentagem de linhas com uma velocidade maior, permitindo aos seus usuários ter viagens mais rápidas em virtude da instalação das faixas exclusivas.

6. CONSIDERAÇÕES FINAIS

Essa pesquisa apresentou uma aplicação de uma nova metodologia de análise de dados de GPS que permite extrair os tempos de viagem e velocidades de linhas que operam em um contexto urbano. Foi possível comparar as velocidades de 308 linhas que operam na cidade de São Paulo, e os resultados mostraram que houve melhora nas velocidades e

redução dos tempos de viagem quando se considere o conjunto de linhas beneficiadas pelas faixas exclusivas como um todo. Os valores médios de aumento de velocidade correspondente ao tempo de viagem percentil 50 foi de 9,13% para as linhas beneficiadas e 0,39% para as não beneficiadas. Vale destacar que as melhorias são diferentes linha a linha, mas a maior parte das linhas beneficiadas por faixas exclusivas teve um aumento de velocidades e redução dos tempos de viagem até mesmo nos dias críticos, representados pelo percentil 90 dos tempos de viagem. Também foi possível concluir que, para algumas linhas, embora tenha havido um aumento na mediana das velocidades das viagens, houve uma redução na velocidade nos casos de trânsito extremo.

Esta análise permitiu ainda demonstrar que os dados de localização de veículos utilizando GPS, quando disponíveis, abrem uma oportunidade valiosa para se avaliarem medidas operacionais tais como a implantação de faixas exclusivas de ônibus, e outras de engenharia de tráfego (por exemplo, sincronização ou prioridade semafórica) de maneira mais simples, rápida, expedita e precisa que no passado, quando eram necessárias pesquisas de campo, mais custosas e menos precisas e confiáveis.

É importante destacar ainda que um dos objetivos da pesquisa é avaliar o potencial de uma metodologia que utilize dados que podem ser facilmente disponibilizados e utilizados de localização de pontos dos trajetos dos veículos a partir de dispositivos de GPS em intervalos curtos, sem necessidade de levantamentos de campo, que são caros e demorados. Alguns aspectos particulares do ambiente de cada

faixa exclusiva, tais como eventuais intervenções no sistema viário, por exemplo nas mãos de direção, na programação de semáforos, não podem ser identificadas através dessa análise proposta. Por outro lado, uma análise mais abrangente, envolvendo centenas de linhas, pode possibilitar identificar eventuais resultados muito discrepantes, que podem sugerir a existência de outros fatores exógenos, como os acima citados, como causa de alterações nas velocidades dos ônibus, permitindo a complementação da análise através de pesquisas de campo pontuais. Esse é um tópico para uma possível investigação futura.

Como pesquisa futura, e de maior complexidade, pode-se sugerir ainda não apenas comparar o ganho dos tempos para os trajetos das linhas inteiras, mas sim, individualmente para cada usuário, verificar os trechos que cada um deles percorre nas linhas do transporte coletivo municipal, e calcular os ganhos respectivos médios em seus deslocamentos diários. Para isso é necessário um modelo de inferência de localização de embarques e desembarques, para a construção dos trajetos dos usuários no sistema de transporte público.

A questão da variabilidade e confiabilidade dos tempos de viagem é uma questão que merece uma análise mais aprofundada em pesquisas futuras. Entre os diversos aspectos que podem ser considerados, sugere-se, por exemplo, o cálculo dos indicadores de confiabilidade dos tempos de viagem e a sua evolução ao longo do tempo. Adicionalmente, pode ser de interesse pesquisar como a variabilidade dos tempos de viagem nos ônibus pode afetar a escolha do modo de transporte pelo usuário, já que quando a variabilidade dos tempos de viagem é reduzida o usuário percebe que não é necessário sair com mais antecedência para que ele assegure chegar ao destino sempre no horário previsto.

Por fim, outra sugestão de pesquisa futura é utilizar essa base de dados gerada ao analisar todas as linhas da cidade para poder criar ferramentas que sugerem a melhor rota ou itinerário para consultas de viagens utilizando transporte público de modo a levar em consideração dados reais de tempo de espera (calculado usando a contagem de viagens completas detectadas na hora), tempo de viagem (pelo percentil 50) e indicadores de confiabilidade. Assim, será possível sugerir a melhor alternativa de rota e itinerário por transporte público em deslocamentos urbanos para cidades que possuam estes dados de GPS disponíveis.

AGRADECIMENTOS

Os autores agradecem à CAPES e ao CNPq pelo apoio à pesquisa, e à SPTrans pela disponibilização dos dados para análise.

REFERÊNCIAS

André, M., & Hammarstöm U. (2000). Driving speeds in Europe for pollutant emissions estimation. *Transportation Research Part D: Transport and Environment*, 5(5), 321-335. DOI: 10.1016/S1361-9209(00)00002-X

CET-SP. (2015). Companhia de Engenharia de Tráfego de São Paulo. Faixas Exclusivas 2014, avaliação de desempenho das faixas implementadas em 2014. Disponível em:

<http://www.cetsp.com.br/media/372687/jtatto_fxexclusivas.pdf>
> Acesso em: 24 de fevereiro de 2016.

Chakroborty, P., & Kikuchi, S. (2004). Using bus travel time data to estimate travel times on urban corridors. *Transportation Research Record: Journal of the Transportation Research Board*, (1870), 18-25. DOI: 10.3141/1870-03.

Cortés, C. E., Gibson, J., Gschwender, A., Munizaga, M., & Zúñiga, M. (2011). Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research Part C: Emerging Technologies*, 19(4), 695-707. DOI: 10.1016/j.trc.2010.12.008

DETRAN-SP. (2016). Secretaria de Planejamento e Gestão. Departamento Estadual de Trânsito de São Paulo. Frota de Veículos em SP – por tipo de veículo para o ano de 2015. Disponível em:
<<http://www.detran.sp.gov.br/wps/portal/portaldetran/detran/estatisticatransito/sa-frotaveiculos>>. Acesso em: 24 de fevereiro de 2016.

FHWA. (2016). Federal Highway Administration. Travel Time Reliability: Making It There On Time, All The Time. Disponível em
<http://ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm>. Acesso em: 09 de março de 2016.

Folha de São Paulo. (2014). São Paulo registra o maior congestionamento da história. Disponível em:<
<http://www1.folha.uol.com.br/cotidiano/2014/05/1459269-sao-paulo-registra-o-maior-congestionamento-da-historia.shtml>>. Acesso em: 24 de fevereiro de 2016.

Google Transit. (2015). Google Transit, documento de referência para desenvolvedores. Disponível em:
<<https://developers.google.com/transit/gtfs/reference>>. Acesso em: 29 de março de 2015.

Hennessy, D. A., & Wiesenthal, D. L. (1997). The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours. *Ergonomics*, 40(3), 348-361. DOI: 10.1080/001401397188198

IBGE. (2016). IBGE, Cidades@: São Paulo. Disponível em
<<http://cidades.ibge.gov.br/xtras/perfil.php?lang=&codmun=355030&search=||infogr%E1ficos:-informa%E7%F5es-completas>>. Acesso em: 24 de fevereiro de 2016.

Mazloumi, E., Currie, G., & Rose, G. (2009). Using GPS data to gain insight into public transport travel time variability. *Journal of Transportation Engineering*, 136(7), 623-631. DOI: 10.1061/(ASCE)TE.1943-5436.0000126.

Numpy. (2016). Disponível em <<http://www.numpy.org/>>. Acesso em 24 de fevereiro de 2016.

Pandas. (2016). Python Data Analysis Library. Disponível em <<http://pandas.pydata.org/>>. Acesso em 24 de fevereiro de 2016.

PMSP. (2016). Prefeitura Municipal de São Paulo. Indicadores. Passageiros Transportados. Disponível em:
<http://www.sptrans.com.br/indicadores/historico_passageiros-transportados.aspx> Acesso em 24 de fevereiro de 2016.

Prefeitura de São Paulo (2016). Prefeitura Municipal de São Paulo. Secretaria Executiva de Comunicação. Aplicativos para transporte coletivo são premiados em Hackatona do Ônibus. Disponível em <

<http://www.prefeitura.sp.gov.br/cidade/secretarias/comunicacao/noticias/?p=160029>>. Acesso em 09 de março de 2016.

Python. (2016). Disponível em <<https://www.python.org/>>. Acesso em 24 de fevereiro de 2016.

SMT. (2016). Prefeitura Municipal de Transportes, Secretaria Municipal de Transportes. Plano de Mobilidade Urbana. Texto Final. Disponível em: <<http://www.prefeitura.sp.gov.br/cidade/secretarias/transportes/planmob/>>. Acesso em: 24 de fevereiro de 2016.

SPTrans. (2015a), São Paulo Transportes. Indicadores. Disponível em: <<http://www.sptrans.com.br/indicadores/>>. Acesso em: 29 de março de 2015.

SPTrans. (2015b), São Paulo Transportes. Área de Desenvolvedores. Disponível em: <<http://www.sptrans.com.br/desenvolvedores/>>. Acesso em: 29 de março de 2015.

Tétreault, P. R., & El-Geneidy, A. M. (2010). Estimating bus run times for new limited-stop service using archived AVL and APC data. *Transportation Research Part A: Policy and Practice*, 44(6), 390-402. DOI: 10.1016/j.tra.2010.03.009.

Uno, N., Kurauchi, F., Tamura, H., & Iida, Y. (2009). Using bus probe data for analysis of travel time variability. *Journal of Intelligent Transportation Systems*, 13(1), 2-15. DOI: 10.1080/15472450802644439

TOMTOM, (2014) Tomtom Americas Traffic Index, 4º. Índice de Tráfego Global, Pesquisa de Tráfego. Disponível em <<http://www.tomtom.com/lib/doc/pdf/2014-05-14%20TomTomTrafficIndex2013annualAme-km.pdf>> e <<http://www.tomtom.com/news/category.php?ID=4&NID=1540&Lid=30>>. Acesso em 1º de setembro de 2015.

Weisbrod, G. Vary, D., & Treyz, G. (2001). Economic implications of congestion. National Cooperative Highway Research Program, Transportation Research Board. Washington, D.C.

Ye, Pengyao et al. (2016) Use of GPS Data to Analyze Urban Bus Travel Time Variability In China. *Transportation Research Board 95th Annual Meeting* (No. 16-5114).