

# Analizando a variabilidade de estimativas de acessibilidade por transporte público a partir de dados de GPS

*Analyzing the variability of public transit accessibility estimates using GPS data*

Carlos Kauê Vieira Braga<sup>1</sup>, Carlos Felipe Grangeiro Loureiro<sup>2</sup>, Rafael H. M. Pereira<sup>3</sup>

<sup>1</sup>Universidade Federal do Ceará, Ceará – Brasil, kaue@det.ufc.br

<sup>2</sup>Universidade Federal do Ceará, Ceará – Brasil, felipe@det.ufc.br

<sup>3</sup>Instituto de Pesquisa Econômica Aplicada – Brasil, rafael.pereira@ipea.gov.br

## Recebido:

16 de setembro de 2019

## Aceito para publicação:

1 de outubro de 2020

## Publicado:

15 de dezembro de 2020

## Editor de área:

Helena Beatriz Cybis

## Palavras-chaves:

Indicadores de acessibilidade.

Dados de GTFS.

Dados de GPS.

Variabilidade no tempo de viagem.

## Keywords:

Accessibility indicators.

GTFS data.

GPS data.

Travel time variability.

DOI:10.14295/transportes.v28i5.2175



## RESUMO

Acessibilidade é um conceito chave no planejamento integrado dos transportes e uso do solo. Dados de *General Transit Feed Specification* (GTFS) têm sido usados com sucesso nos últimos anos como insumo na mensuração dos níveis de acessibilidade, mas são baseadas em viagens programadas de transporte público para estimar tempos de viagem entre pares origem-destino. Por outro lado, dados de GPS têm sido aplicados na atualização de tabelas horárias de viagens realizadas, embora com limitações metodológicas e computacionais, baseando-se em medidas de tendência central dos tempos de viagem para recompor o arquivo GTFS. Usando um mês de dados de GPS dos ônibus em Fortaleza, analisou-se a variabilidade da acessibilidade a partir dos tempos de viagem observados, calculados considerando sua dispersão. Os resultados demonstram que há variabilidade significativa na estimação dos indicadores de acessibilidade, tanto no espaço como entre atividades de trabalho e educação, podendo impactar na avaliação de intervenções no sistema de transportes.

## ABSTRACT

Accessibility is a key concept in transport and land use integrated planning. General Transit Feed Specification (GTFS) data have been successfully used in recent years as input to measure accessibility levels, but it is based on scheduled transit trips to estimate travel times between origin-destination pairs. On the other side, GPS data have been applied to update actual trip time tables, although with methodological and computational limitations, based on travel times' measures of central tendency to reconstruct the GTFS file. Using one month of GPS bus data in Fortaleza, this work analyzed the variability of accessibility through empirical travel times, considering its dispersion. The results show that there is a significant variability in the estimation of the accessibility indicators, both in space and between work and school activities, possibly affecting the assessment of interventions on the transportation system.

## 1. INTRODUÇÃO

A acessibilidade, definida como o potencial de oportunidade de interação no espaço urbano (Hansen, 1959), é um conceito chave para o planejamento das cidades. Geurs e van Wee (2004) entendem acessibilidade às atividades como um importante instrumento que permite ao planejador avaliar os impactos que políticas e intervenções nos sistemas de uso do solo e de transportes têm no funcionamento da sociedade.

Até poucos anos atrás, a mensuração da acessibilidade era limitada pela falta de dados e pelo baixo poder computacional. Recentemente, a expansão da capacidade computacional e a disponibilidade de dados padronizados da oferta de transporte público, como o *General Transit Feed Specification* (GTFS), fomentaram a criação de ferramentas dedicadas que analisam e aplicam algoritmos de escolha de rota sobre a rede (módulos dedicados do ArcGis e *OpenTripPlanner*, por exemplo). Esses desenvolvimentos recentes facilitaram a estimação de matrizes de tempos de viagem entre pares origem-destino que subsidiam o cálculo de indicadores de acessibilidade por transporte público e têm sido crescentemente utilizados na literatura (Owen e Levinson, 2015; Mayaud *et al.*, 2018; Pereira *et al.*, 2019). Via de regra, no entanto, tais análises de acessibilidade têm se baseado em dados de GTFS que trazem as informações de tempos de viagens planejadas dos veículos, ignorando como a variabilidade da performance do sistema de transporte público ao longo do dia pode afetar a acessibilidade proporcionada pelo sistema.

Dados de GPS, por outro lado, podem oferecer informações de localização de toda a frota de um sistema de transporte público. Utilizando-se de métodos computacionais e de geoprocessamento, essas localizações podem ser traduzidas em tempos de deslocamento entre paradas do sistema e transformadas em um formato capaz de corrigir as imprecisões de tempo de viagem dos dados de GTFS. Trabalhos anteriores de Wessel *et al.* (2017) e de Arbex *et al.* (2016) realizaram essa transformação, contando com todos os dados de localização da frota de Toronto e São Paulo, respectivamente.

Nesses estudos, a utilização de dados observados da frota para esse fim ocorreu com amostras de mais de um dia de GPS para corrigir os horários programados, mas não utilizaram tal informação para calcular tempos de viagem levando em conta a dispersão dos mesmos (considerou-se somente a média dos tempos em cada trecho) (Wessel *et al.*, 2017; Arbex *et al.*, 2016). Num contexto de avaliação de políticas públicas, a utilização somente de medidas de tendência central dos tempos de viagem na rede pode levar à superestimação das oportunidades que são acessadas, principalmente em sistemas de transporte público que têm alta variabilidade de nível de serviço. Entende-se que a incorporação de medidas de dispersão dos tempos de viagem observados pode levar a uma maior robustez no uso da acessibilidade para avaliar alternativas.

Portanto, este trabalho tem como objetivo geral analisar a variabilidade da acessibilidade para oportunidades de trabalho e educação, utilizando informações de tempos de viagem e sua dispersão, a partir de dados de GPS de localização da frota. Esse esforço analítico se dá em um contexto da avaliação de intervenções no sistema de transportes baseado em métricas de acessibilidade. Como objetivo específico, busca-se propor um método para a reconstrução de tabelas de horários programados do GTFS a partir de dados de GPS, capaz de captar a variabilidade na performance do sistema de transporte público durante um período de dados arquivados de localização da sua frota. Este trabalho utiliza ainda o sistema de transporte público da cidade de Fortaleza (Brasil) como estudo de caso, para demonstrar como a metodologia proposta permite captar diferenças entre as estimativas de acessibilidade baseadas nos dados programados de GTFS e em dados observados a partir de GPS, incorporando sua dispersão.

Na seção 2, faz-se uma revisão da literatura sobre a temática, apresentando o estado da arte na estimação da acessibilidade com os dados de GTFS. A seção 3 descreve a metodologia proposta para a reconstrução da tabela de horários (*stop\_times.txt*) e o método para análise dos indicadores de acessibilidade. Na seção 4, aplica-se essa metodologia para o caso da cidade de Fortaleza. Por fim, a seção 5 apresenta as principais conclusões acerca do método estabelecido e dos resultados encontrados.

## 2. REVISÃO DA LITERATURA

O conceito clássico de Hansen (1959) define acessibilidade como o potencial de oportunidade de interação no espaço urbano. Em uma revisão mais recente, Geurs e van Wee (2004) entendem a acessibilidade como as oportunidades para indivíduos participarem em atividades no espaço urbano, sendo também utilizada para medir o impacto que políticas e intervenções nos sistemas de uso do solo e transportes têm no funcionamento da sociedade.

Até a metade da década de 2000, a estimação de indicadores de acessibilidade por transporte público se baseava em informações simplificadas das redes de transportes (Owen e Levinson, 2015). Os métodos utilizados até então não conseguiam incorporar etapas importantes do deslocamento, como o tempo de espera inicial e o tempo de espera por integrações (Lei e Church, 2010). A criação em 2005 do formato de arquivo GTFS (*General Transit Feed Specification*) abriu novas possibilidades para o uso de informação da oferta de transporte público no planejamento de transportes. Esse formato é composto de diversos arquivos *.txt* que delimitam informações sobre a oferta de transporte público de uma cidade, permitindo ao órgão gestor consolidar todas as informações da oferta do sistema e divulgar o arquivo *.zip* com todo o seu conteúdo. A sua implementação por agências de transportes, entretanto, requer uma estruturação de dados e pessoal capacitado que nem sempre está disponível em cidades de médio e pequeno porte.

A sua adoção como formato padrão de consolidação de dados de transporte público por milhares de agências de transportes do mundo possibilitou a criação e padronização de ferramentas para sua análise, com destaque para ferramentas de escolha de rota. As duas principais ferramentas (*ArcGIS* e *OpenTripPlanner*) tomam como input um arquivo de GTFS e a rede viária da cidade para a construção de um objeto que represente a cidade como uma *network*, onde as paradas de ônibus atuam como nós, os horários agendados entre paradas atuam como *edges*, e o tempo de viagem entre paradas atua como o custo (Wessel *et al.*, 2017). Dessa forma, aplicam um algoritmo de caminho mínimo e retornam como resultado o tempo de viagem entre pontos desejados, considerando as etapas de caminhada, espera, integração e tempo dentro de veículo. Com os dados de GTFS e ferramenta disponíveis, diversos estudos recentes estimaram tempos de viagem entre pares origem-destino para subsidiar o cálculo de indicadores de acessibilidade por transporte público (Owen e Levinson, 2015; Wessel *et al.*, 2017; Mayaud *et al.*, 2018; Pereira, 2019; Mavoa *et al.*, 2012; Farber e Fu, 2017; Stępnia *et al.*, 2019).

A utilização dos dados de GTFS para esse fim, entretanto, ainda apresenta limitações importantes. Os estudos analisados acima assumem como uma premissa implícita a de que os horários programados fornecidos pela agência de transportes por meio do GTFS são confiáveis. É sabido, contudo, que horários programados podem diferir significativamente dos horários reais (Mandelzys e Hellings, 2010).

Recentemente, estudos vêm tentando incorporar dados de localização da frota na melhoria da estimação de indicadores de acessibilidade por transporte público. Os métodos desses trabalhos buscam transformar dados brutos observados de localização da frota em uma tabela de horários dos veículos no formato *stop\_times.txt* do GTFS, permitindo assim estimar com maior precisão a acessibilidade. Dentre esses, destaca-se o esforço de Wessel *et al.* (2017), propondo uma metodologia que utiliza dados de GPS para corrigir os dados de GTFS em Toronto (Canadá). Os dados de GPS são disponibilizados em tempo real através de uma API, que automaticamente detecta a linha, o sentido, e quando há o fim de uma viagem e começo de outra. O algoritmo desenvolvido pelos autores coletou dois dias de registros e buscou, inicialmente, detectar erros nessas informações. Em seguida, os autores estabeleceram um *buffer* de 20 metros em relação

a cada parada (coletada a partir do arquivo *stops.txt* do GTFS) para estabelecer o ponto de GPS mais próximo, estimando os tempos de parada de cada viagem. Com a estimação dos tempos de parada em cada viagem realizada (e já com as informações de linha, sentido e veículo), o novo arquivo fica análogo ao arquivo agendado *stop\_times.txt*, sendo utilizado para o roteamento pelo *OpenTripPlanner*, assim como na estimação do tempo de viagem entre pares origem-destino (no caso, centróides de setores censitários).

Wessel e Farber (2019) aplicaram essa mesma metodologia para quatro cidades da América do Norte com o objetivo de calcular uma gama de indicadores de acessibilidade para empregos antes e depois da correção, analisando a variabilidade da acessibilidade quando feita a comparação entre o GTFS programado e o corrigido pelo GPS. Os autores encontraram que as estimativas de acessibilidade variam em média entre 5% a 15% (ou até mais); apontam ainda que essa variação apresenta dependência espacial, ocorrendo de maneira mais sistemática em certas regiões das cidades. Por fim, argumentam que utilizar acessibilidade baseada em GTFS programado seria como utilizar acessibilidade para carros sem levar em conta congestionamentos, também tendo grande impacto em estudos que comparem acessibilidade entre modos ou que utilizem esse indicador como medida de desempenho para avaliar alternativas de intervenção.

Noutro estudo, Arbex *et al.* (2016) também utilizaram dados de GPS para corrigir os arquivos de GTFS para o sistema de transporte público de São Paulo e analisar a acessibilidade de cinco grandes favelas na cidade. A metodologia de transformação dos dados de GPS em dados *stop\_times.txt* de GTFS foi detalhada em outro trabalho (Arbex e da Cunha, 2016). Nesse último trabalho, os autores propuseram uma metodologia para agregar tempos de viagens entre paradas estimados a partir de um grande período de dados arquivados de GPS. Não foi realizada uma análise da variabilidade da acessibilidade após a incorporação dos tempos de viagem observados.

Embora esses estudos recentes demonstrem a importância da utilização dos dados de tempo de viagem observados no cálculo de indicadores de acessibilidade, há ainda uma relevante lacuna na literatura no que diz respeito à utilização da amostra advinda dos dados de GPS. Os trabalhos encontrados utilizaram amostras de mais de um dia de localização da frota, mas no fim se restringiram a considerar sua tendência central fazendo a média dos valores amostrados. Wessel e Farber (2019) tiveram acesso à localização de todos os veículos em cada uma das cidades analisadas, por cinco dias, com exceção de algumas linhas onde foram usados os dados programados. Então, construíram um GTFS para cada dia e calcularam a acessibilidade resultante do serviço observado no dia correspondente; computando, por fim, uma acessibilidade média dessa amostra. Já Arbex *et al.* (2016) aplicaram seu algoritmo para 20 dias úteis, não sendo explicitado como os tempos estimados entre paradas foram agregados para um arquivo consolidado de *stop\_times.txt*.

É entendido que as amostras coletadas podem ser utilizadas para além de uma análise baseada em valores médios do sistema, principalmente em realidades nas quais o tempo de viagem por transporte público pode ser tão variável ao longo dos dias. Por exemplo, Mazloumi *et al.* (2010) utilizaram uma grande amostra de dados de GPS para estudar a variabilidade do tempo de viagem de uma rota com o uso de percentis. Ainda falta uma abordagem desse tipo para o estudo da acessibilidade.

Ademais, a análise do impacto da utilização dos dados de tempo de viagem observados no trabalho de Wessel e Farber (2019) foi feita somente para atividades de trabalho, que tendem a ter uma distribuição espacial mais concentrada. Entende-se, entretanto, que outras atividades

(como educação) podem se distribuir de forma distinta no espaço urbano, levando a resultados diferentes de variabilidade da sua acessibilidade. É identificada então uma lacuna adicional no tipo de atividade para a qual é feita a análise dos indicadores de acessibilidade.

Dadas as limitações e lacunas descritas acima, o método desenvolvido a seguir busca avançar nessas temáticas.

### 3. PROPOSTA METODOLÓGICA

Inicialmente, propõe-se a reconstrução dos arquivos *stop\_times.txt* a partir de tempos de viagens coletados dos dados de GPS, tanto para uma medida de tendência central como para uma medida de dispersão da variável aleatória tempo de viagem. Por fim, é proposto o método para o cálculo dos indicadores e análise da variabilidade da acessibilidade para os diferentes cenários.

#### 3.1. Consolidação e integração dos dados

Os dados brutos de GPS do Sistema Integrado de Transporte Público de Fortaleza (SIT-FOR) apresentam colunas com a identificação do veículo, o momento do registro (aqui também chamado de “hora”) e as coordenadas. A informação da linha do veículo, que é necessária para o prosseguimento com o método, não é disponibilizada. Para isso, são utilizados os dados de bilhetagem, que apresentam a identificação dos veículos e a linha correspondente.

Entretanto, essa identificação dos veículos é diferente nas bases. Existe um dicionário que faz a ligação dessas duas identificações, mas é conhecido que essa junção entre as bases não é perfeita. Isso significa que vários veículos da base do GPS não terão suas linhas estimadas, fazendo com que estes sejam excluídos das etapas seguintes. Os impactos serão discutidos nos resultados.

#### 3.2. Transformação de localizações para uma tabela de horários

Esta etapa busca transformar pontos de GPS para um formato de uma tabela de horários de passagem em cada parada semelhante ao *stop\_times.txt*. Isso é alcançado através de quatro passos: (i) filtro espacial; (ii) definição das viagens; (iii) definição do sentido das viagens; e (iv) estimativa do momento de passagem em cada parada.

A informação da linha trazida da etapa anterior ainda pode apresentar inconsistências. Primeiro, um mesmo veículo pode servir a mais de uma linha no mesmo dia. Segundo, pontos fora da rota em trajetos de ida/volta da garagem ainda estão presentes. Para isso, é proposto um filtro espacial, onde é feito um *buffer* de 300 metros em relação ao *shape* de cada linha (*shape* de ida e volta, juntos, advindos do arquivo *shapes.txt* do GTFS), sendo mantidos somente os pontos que estão inseridos dentro dessa área e no começo/fim da viagem. Com essa informação, o fim de uma viagem é determinado no primeiro registro de GPS dentro da área de começo/fim. O início da viagem seguinte acontece no último registro antes do veículo sair da área de início/fim de viagem.

Para definir o sentido da viagem (ida ou volta), inicialmente os registros de cada viagem são ordenados pelo momento do registro. Então, é feito um *snap* de cada registro de GPS para a parada mais próxima de ida daquela linha, tendo essas paradas uma identificação com a sequência que elas acontecem (*stop\_sequence*). Se os pontos de GPS daquela viagem estiverem indo no mesmo sentido da sequência de paradas, a viagem estava indo no sentido da ida; senão, será viagem de volta.

Por fim, para recompor a tabela de horários dos veículos, é necessário estimar o momento que cada veículo passou pelas paradas da sua correspondente linha e sentido. A abordagem aqui adotada leva em conta a distância acumulada percorrida pelos registros de GPS e a distância acumulada da posição das paradas, na linha em questão. Primeiro, é feito o *snap* do ponto de GPS para a linha. Segundo, é garantido que o primeiro ponto de GPS e a primeira parada estejam no mesmo ponto. A partir desse ponto, então, são calculadas as distâncias acumuladas na linha, tanto dos pontos de GPS como das paradas. Do lado dos dados de GPS, cada ponto agora tem a distância acumulada na rede e o momento do registro de localização. É então feita uma interpolação linear para estimar a hora (momento) em cada distância acumulada da parada.

### 3.3. Agregação dos tempos de viagem por trecho e intervalo

A etapa anterior tem como resultado uma tabela de horários (uma para cada dia) de passagem em cada parada de todos os veículos captados pelo GPS, com seus dados sendo consolidados e integrados. Tendo em vista que o objetivo é analisar a variabilidade em virtude da dispersão do tempo de viagem, é então proposta uma agregação dos tempos de viagem dos veículos por trecho e intervalo de 15 minutos.

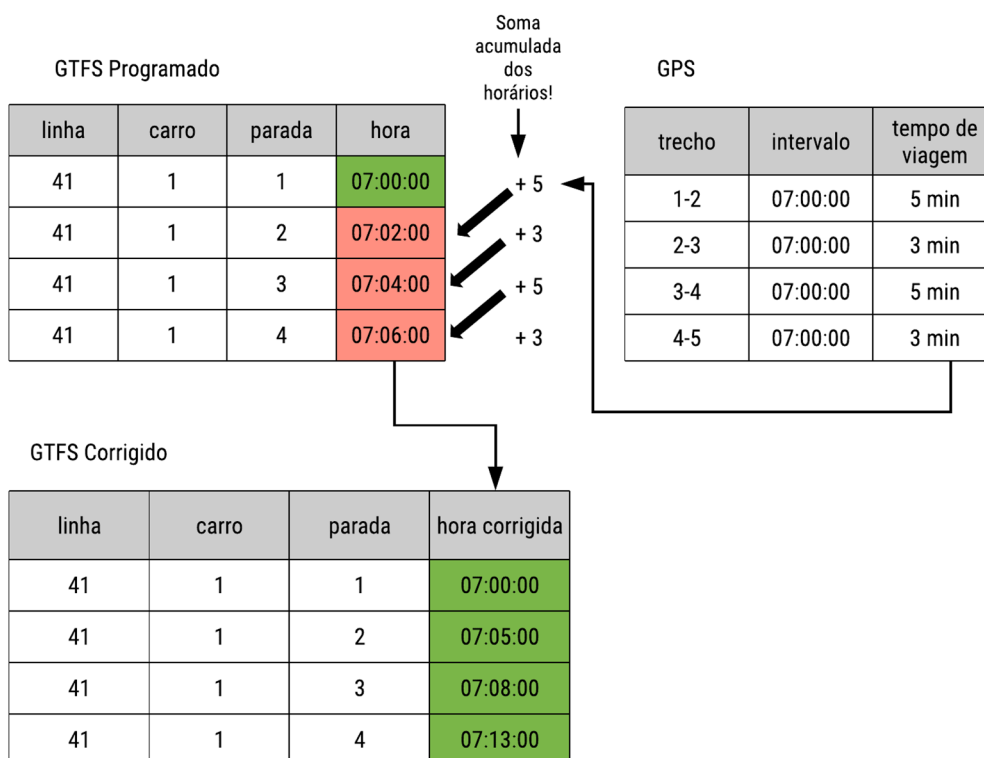
O trecho é entendido como a unidade espacial básica de um arquivo *stop\_times.txt*. Eles são definidos como qualquer segmento entre duas paradas onde cada veículo estava programado para trafegar. A agregação aqui proposta calcula o tempo que cada veículo levou para percorrer cada trecho ao longo de todo o dia, agregando-os para estimar como esses tempos de viagem em cada trecho variam dentro de intervalos de 15 minutos. Cada trecho e intervalo, portanto, terá uma distribuição de tempos de viagem coletados a partir de vários dias de dados de localização. Buscando representar um tempo de viagem central e um tempo de viagem que incorpore a dispersão em cada um desses trechos e intervalos, são calculados a mediana e o 85º percentil dos valores amostrais. Para aumentar a confiabilidade das medidas, os tempos de viagem das combinações de trecho e intervalo que tiverem menos que 10 observações são substituídos pelo tempo de viagem já disponível no GTFS programado.

Acredita-se que essa abordagem avance na realizada por Arbex *et al.* (2016), sendo ainda distinta da proposta do trabalho de Wessel *et al.* (2017). O primeiro também fez uma agregação por trechos a partir de dados arquivados de GPS, mas não relatou a resolução temporal de agregação (se foi por 15 minutos, ou por hora, ou por hora pico) e a medida estatística de agregação utilizada (se foi média ou mediana). O último não propôs uma agregação dos tempos, calculando a acessibilidade diretamente para os horários que foram observados.

### 3.4. Incorporação dos tempos de viagens agregados no *stop\_times.txt*

O resultado da etapa anterior gera os valores da mediana e do 85º percentil dos tempos de viagem entre cada parada (trecho) para cada intervalo de 15 minutos. O método de reconstrução busca substituir os tempos de viagem entre paradas que foram programados (presentes no *stop\_times.txt*) com esses tempos de viagem observados, para intervalos de 15 minutos iguais, depois somando-os cumulativamente. Para isso, é necessário estabelecer duas premissas sobre o sistema de transporte público de Fortaleza que estão traduzidas nos dados de GTFS: 1) o primeiro horário de partida do veículo no dia é correto; e 2) a quantidade de veículos rodantes é correta. De acordo com os gestores do sistema, ambas as premissas são factíveis.

A Figura 1 esquematiza o método desenvolvido para reconstruir o arquivo com os horários programados. A primeira tabela à esquerda apresenta os horários agendados de um carro de uma linha 41, nas primeiras 4 paradas. Na direita, tem-se a tabela agregada do GPS com o tempo de viagem mediano de todos os carros que passaram em cada um dos trechos determinados, no intervalo de 07:00:00 (valores são agregados para o intervalo de 15 minutos mais próximo). Confirmado que os intervalos da hora programada e do GPS são condizentes, é assumido então que o horário da primeira partida do carro no dia é o horário correto (em verde). Dessa forma, para cada trecho de viagem correspondente, é feita a soma acumulada dos tempos de viagens a partir do primeiro horário, resultando na tabela (na parte inferior) o horário acumulado corrigido.



**Figura 1:** Método de reconstrução de arquivos *stop\_times.txt* a partir de tempos de viagem agregados por trechos e intervalos

A soma acumulada dos horários corrigidos é feita até o fim de cada viagem. Ao chegar o fim, outra medida estimada a partir dos dados de GPS é utilizada: o tempo entre viagens. Para cada linha, para o intervalo de 30 minutos mais próximo, é estimado o tempo entre uma viagem e outra (com a distinção se é uma transição entre ida e volta ou volta e ida). Isso é importante principalmente para linhas que utilizam terminais de integração, onde o tempo decorrido desde as manobras até o desembarque e o embarque dos passageiros pode levar minutos, com esse tempo de atraso não sendo explicitamente incorporado no *stop\_times.txt* programado.

Essa correção do tempo entre viagens é feita até que seja encontrado um intervalo entre viagens programadas entre 30 e 59 minutos. Esse intervalo foi identificado e confirmado como uma pausa programada para os funcionários do veículo, sendo normalmente respeitado. Então, quando for encontrada uma pausa entre viagens, o valor desse intervalo é incorporado e somado à hora corrigida.

Esse passo a passo detalhado nos últimos parágrafos é feito para cada veículo na tabela de horários programados do *stop\_times.txt*, utilizando tanto o tempo de viagem mediano como o 85º percentil em cada trecho para fazer a soma acumulada. O produto final são dois arquivos de GTFS: um possuindo uma tabela de horários *stop\_times.txt* corrigido a partir da mediana dos tempos de viagem (P50) e outro corrigido a partir do 85º percentil (P85). Esses novos arquivos de GTFS são chamados de GTFS Empírico P50 e GTFS Empírico P85, respectivamente.

O GTFS Empírico P50 é considerado o contraponto do GTFS Programado. A mediana é escolhida porque constata-se uma assimetria considerável na distribuição dos tempos de viagem nos trechos, o que compromete a representatividade da média como medida de tendência central. Para o valor de dispersão, o 85º percentil dos tempos de viagem nos trechos é escolhido como uma forma de oferecer uma medida que dê mais segurança nas estimativas de acessibilidade. A sua utilização para recompor o *stop\_times.txt* materializada no GTFS Empírico P85 também pode ser interpretada como uma forma de garantir que um usuário consiga acessar as oportunidades da cidade em pelo menos 85% das vezes, assumindo um mesmo horário de partida.

### 3.5. Análise da variabilidade da acessibilidade

A construção dos arquivos GTFS Empírico P50 e P85, junto ao GTFS Programado, fornece os dados de oferta do transporte público que possibilita a análise comparativa da acessibilidade. Essa análise é feita por meio de um indicador cumulativo de oportunidades que estima a quantidade de oportunidades que podem ser alcançadas dado um certo tempo de viagem limite. Esse indicador é uma das medidas de acessibilidade mais amplamente utilizadas por estudos que usam dados de GTFS ou de GPS para análises de acessibilidade de sistemas de transporte público (Wessel *et al.*, 2017; Owen e Levinson, 2015; Pereira, 2019). Uma limitação desse indicador é a premissa de que todas as oportunidades dentro do limite de tempo de viagem estipulado são igualmente desejáveis pelos viajantes. Por exemplo, duas escolas que são acessíveis em menos de 30 minutos terão pesos iguais, mesmo que uma escola esteja a uma distância de 1 km e a outra a uma distância de 5 km.

O primeiro passo para calcular o indicador de acessibilidade é estimar o tempo de viagem entre pares origem-destino. Para isso, é utilizada a ferramenta de roteamento *OpenTripPlanner* (<https://www.opentripplanner.org>), que é *open source* e faz estimativas de tempo de viagem detalhadas considerando todas as etapas do deslocamento de porta-a-porta: o tempo de caminhada até a parada, o tempo de espera, o tempo no veículo e o tempo de espera em possíveis integrações.

Os pares origem-destino são produzidos a partir dos centroides das agregações espaciais H3 (<https://eng.uber.com/h3>) desenvolvidas pelo Uber. A agregação escolhida é um hexágono que tem tamanho de diagonal menor igual a 357 metros, o que possibilita analisar a variação da acessibilidade em alta resolução espacial. Já foi mostrado que esse tipo de agregação é capaz de representar algumas tendências de acessibilidade (como acessibilidade maior perto de corredores de média/alta capacidade) que outras agregações maiores não conseguem (Pereira, 2019).

Para cada um dos GTFS comparados foram estimadas cinco matrizes de tempo de viagem entre cada um dos 2700 pares origem-destino de Fortaleza; cada matriz partindo a cada 15 minutos entre 6h45 e 7h45 da manhã. O cálculo de matrizes para várias horas de partida



é feito porque, para viagens de transporte público, até pequenas diferenças no horário do começo da viagem podem implicar em grandes diferenças no tempo total de viagem, principalmente através da perda de integrações e elevados tempos de espera (Stępniaik *et al.*, 2019). O valor estimado do tempo de viagem entre cada par origem-destino é calculado como a mediana dos tempos nos diferentes horários de partida estimados ( $n=5$ ). O resultado são três matrizes de tempo de viagem, uma para cada GTFS (um com serviço programado e outros dois reconstruídos a partir do GPS, P50 e P85).

São calculados indicadores cumulativos para cada um dos GTFS para atividades de emprego e educação, com tempos limites de 65 e 50 minutos, respectivamente. Os tempos limites são extraídos a partir dos dados de bilhetagem, sendo calculado o tempo que pelo menos 75% da população leva nos seus deslocamentos de hora pico para cada uma das atividades. As atividades de emprego são representadas pelos vínculos formais registrados na Relação Anual de Informações Sociais (RAIS), do Ministério do Trabalho, do ano de 2017. As atividades de educação são advindas do Censo Escolar de 2015, sendo representadas pela quantidade de matrículas para ensino infantil, fundamental e médio de todas as escolas públicas (municipais, estaduais e federais) de Fortaleza.

Tendo os indicadores calculados, é estabelecido então o método de análise comparativa. As comparações, divididas por atividades, são feitas de forma descritiva, analisando de forma espacial (mapas) e não-espacial (*boxplots*) valores absolutos e relativos de variação. Duas comparações são sugeridas: uma entre o GTFS Programado e o GTFS Empírico P50, que busca ver o efeito da utilização dos dados observados da frota na acessibilidade, de acordo com o que tem sido feito na literatura; e outra entre o GTFS Empírico P50 e o GTFS Empírico P85, que propõe a utilização de uma medida de dispersão em comparação com uma medida de tendência central para análises de acessibilidade. A diferença relativa percentual é representada na forma da diferença logarítmica entre cenários,  $\log(\frac{x}{y})$ , que busca eliminar problemas de assimetria da medida tradicional de mudança relativa  $(100(y - x)/x)$  (Tornqvist *et al.*, 1985).

#### 4. RESULTADOS E DISCUSSÕES

Fortaleza é a quinta cidade mais populosa do Brasil, com uma população estimada de 2,7 milhões de habitantes (IBGE, 2019). O seu sistema de transporte público conta com 318 linhas de ônibus e um total de aproximadamente 1 milhão de viagens realizadas diariamente. Para referência, a Figura 2 mostra a distribuição espacial de variáveis demográficas e das atividades que serão consideradas no cálculo dos indicadores de acessibilidade.

A metodologia foi aplicada para dados de GPS de 19 dias úteis do mês de setembro/2018. Como antecipado, a etapa da consolidação e integração dos dados impôs uma exclusão de parte da amostra de GPS. Em média, não foi possível estimar a linha de cerca de 15% dos veículos do sistema (275 veículos de um total de 1800/dia), sendo esses descartados da amostra. É importante destacar que esse descarte não é aleatório: foi observado que grande parte das linhas com prefixo 7 (linhas complementares) não tiveram suas localizações estimadas. Essas linhas servem muitas vezes bairros que não têm uma demanda suficiente para ter uma linha regular. A não inclusão desse serviço nas etapas seguintes pode comprometer as estimativas de tempo de viagem para áreas localizadas em bairros mais afastados. Outro descarte de dados foi realizado na etapa do filtro espacial utilizando os dados de GTFS: foi identificado que 27 linhas do SITFOR não estavam presentes nos dados de GTFS, e isso fez com que pontos de GPS dessas linhas

fossem deletados da amostra. Isso ocasionou um novo descarte de cerca de 5% da amostra dos dados de GPS.

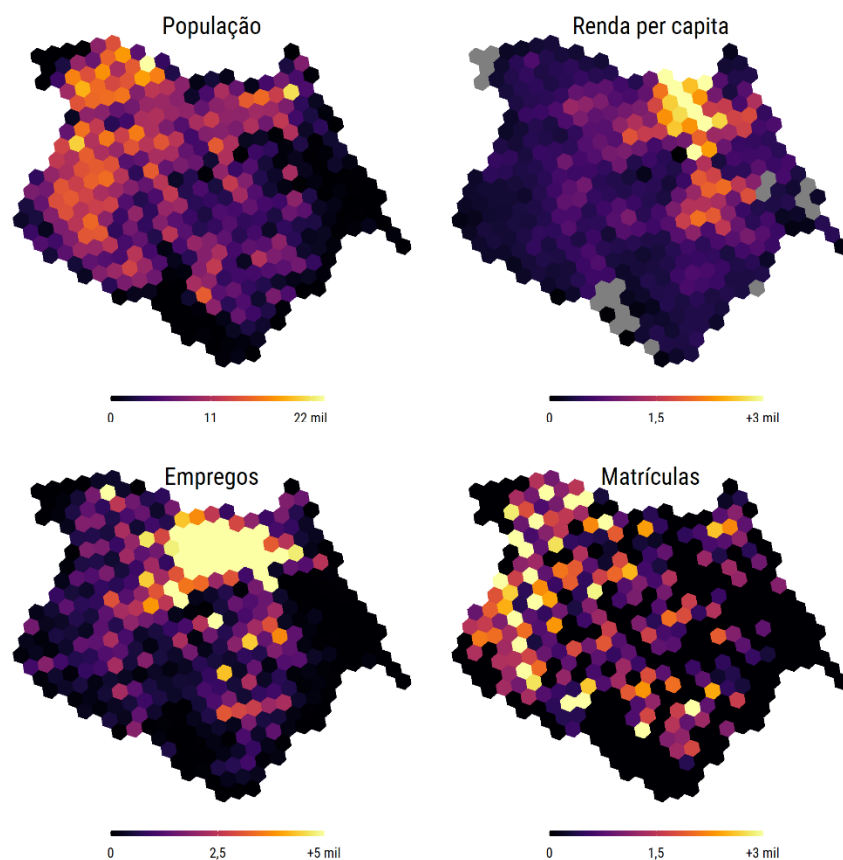


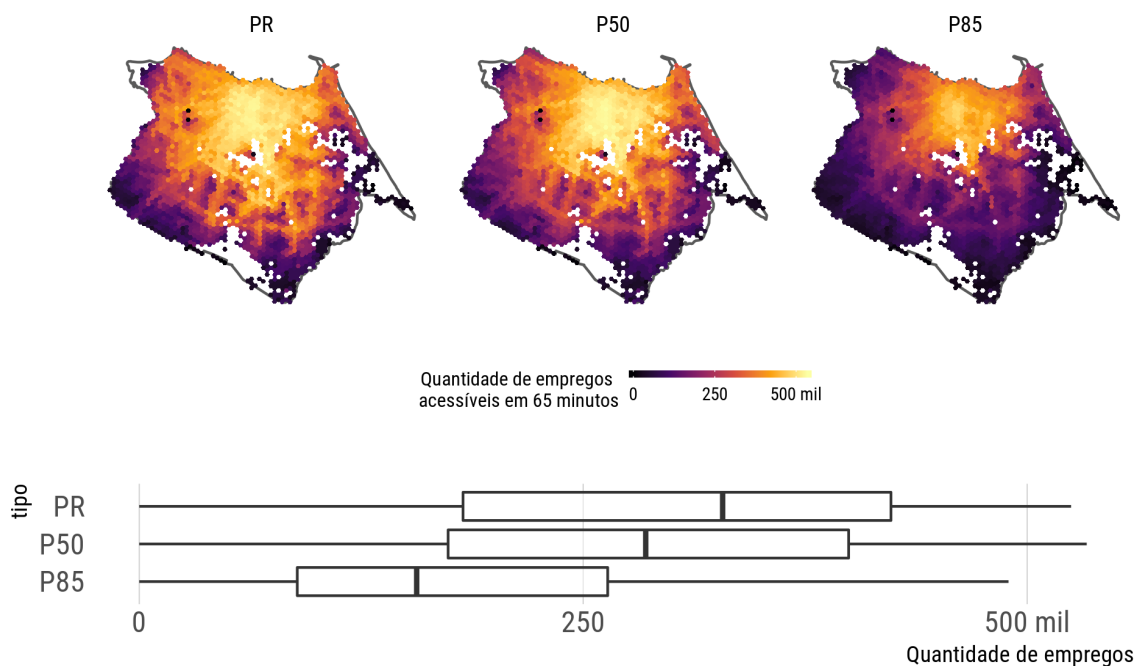
Figura 2. Distribuição espacial dos indicadores demográficos e de uso do solo para Fortaleza

Os dois descartes descritos acima podem impactar na qualidade da amostra resultante do processo de agregação dos tempos de viagem por trecho e intervalo. Esse processo de agregação gerou em torno de 120 mil combinações de trecho entre paradas e intervalo de 15 minutos (5400 trechos x 22 intervalos do período da manhã, aproximadamente). Desse valor, cerca de 17% das combinações tiveram uma amostra menor que 10 tempos de viagem. Para essas combinações o tempo de viagem programado foi utilizado. A maioria dessas combinações são de trechos que estão presentes em áreas mais periféricas e de pouco tráfego de veículos em geral, o que em certa medida torna o descarte menos problemático tendo em vista que são áreas que apresentam menor variabilidade nos tempos de viagem.

#### 4.1. Análise da variabilidade da acessibilidade para atividades de trabalho

A Figura 3 mostra a distribuição espacial (parte superior) e não-espacial (parte inferior) da acessibilidade para oportunidades de empregos dentro de um tempo limite de 65 minutos, com os mapas representando a acessibilidade pelo GTFS Programado (PR), GTFS Empírico P50 e GTFS Empírico P85. Em primeiro lugar, nota-se como a concentração de empregos no centro da cidade, somada a uma maior oferta de transporte público nessa região, leva a uma maior acessibilidade. Segundo, a alta resolução espacial permite visualizar como a acessibilidade é melhor próximo aos principais corredores de ônibus (manchas retas mais claras).

Na questão da variabilidade, fica evidente pelos mapas que a acessibilidade para tempos de viagem P85 é menor que para os demais cenários. Apesar de ser difícil tirar conclusões a partir dos mapas PR x P50, os *boxplots* mostram que os valores de acessibilidade com o GTFS Programado tendem a ser maiores que os do GTFS Empírico P50 - os valores dos percentis 25 e 75 são levemente superiores, enquanto que a mediana aparenta ter uma maior diferença. De fato, a mediana das oportunidades de emprego acessíveis para o PR foi de 330 mil empregos, enquanto que para o P50 foi de 285 mil empregos - uma diferença de 13%. Já a mediana dos empregos acessíveis para o P85 foi de 165 mil empregos. Esses resultados ilustram como o uso de dados de GTFS programado pode levar a uma superestimação da acessibilidade efetivamente vivenciada pela população.

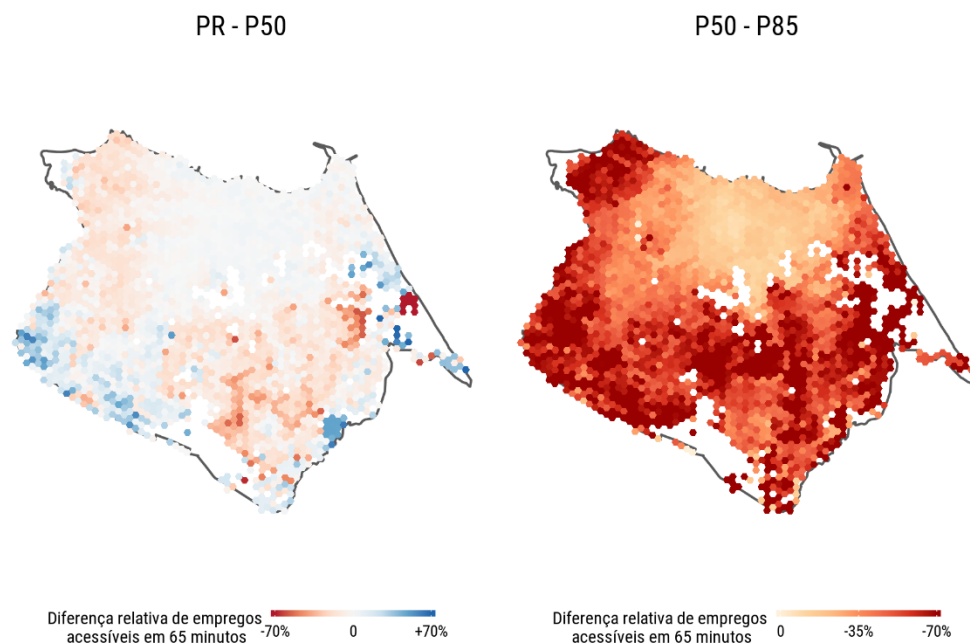


**Figura 3.** Distribuição espacial (acima) e não-espacial (abaixo) dos indicadores de acessibilidade para trabalho para cada um dos GTFS

Para destacar como as diferenças das estimativas de acessibilidade a partir de diferentes fontes variam espacialmente, a Figura 4 apresenta a diferença relativa de acessibilidade a partir dos dados de GTFS Programado (PR) e Empírico P50, assim como a diferença entre as estimativas P50 e P85. Pelo mapa da esquerda é observado que zonas com alta diferença negativa de acessibilidade estão presentes no entorno da área central de Fortaleza. Isso indica que o GTFS Programado falha ao captar prováveis condições de congestionamento nas vias (captadas pelo GTFS Empírico P50) que levam os usuários da periferia para o centro da cidade. O padrão no mapa da direita (P50 - P85) é semelhante ao do primeiro mapa: essas mesmas áreas parecem ser as mais prejudicadas pela dispersão nos tempos de viagens. Constata-se, portanto, que vias nas quais é mais difícil a linha se manter no horário tendem a ter uma variabilidade maior nos tempos de viagem ao longo dos dias.

O mapa da diferença relativa entre P50 e P85 também mostra que o P50 superestima a acessibilidade em mais de 75% para diversas aglomerações localizadas na periferia da cidade. Em certas regiões, esse número pode representar uma superestimação absoluta de mais de 150 mil

empregos, o que significa quase 30% dos empregos totais da cidade. Números de diferença absoluta entre os cenários mostram uma superestimação mediana de 95 mil empregos (17% do total de empregos da cidade), com um 75º percentil de 124 mil (23% do total).



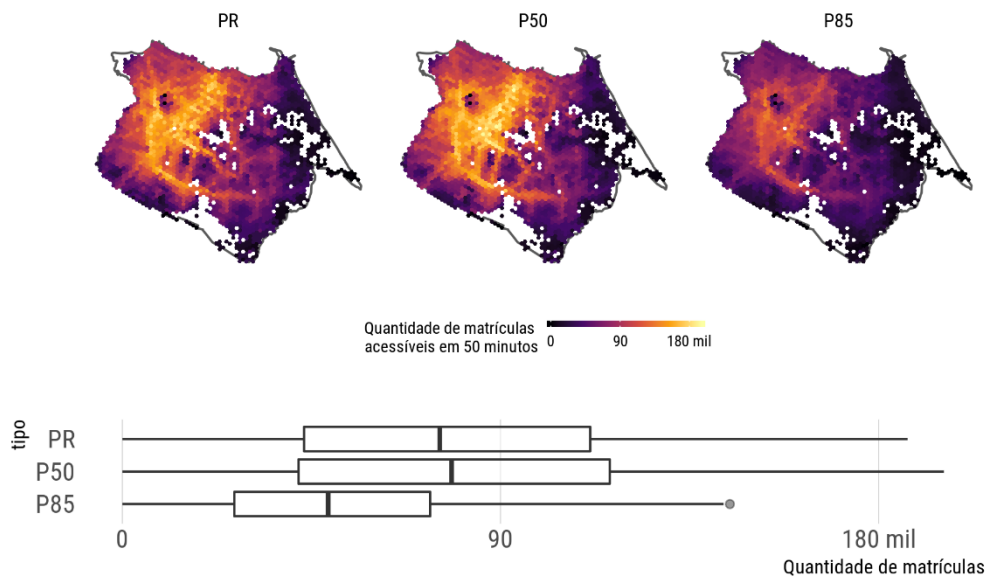
**Figura 4.** Diferença relativa de empregos acessíveis para comparação GTFS Programado x GTFS Empírico P50 (à esquerda) e GTFS Empírico P50 x GTFS Empírico P85 (à direita)

#### 4.2. Análise da variabilidade da acessibilidade para atividades de educação

Numa análise semelhante, mas agora focando no acesso a oportunidades de educação, a Figura 5 apresenta a distribuição da acessibilidade a escolas públicas representadas pela quantidade de matrículas. A distribuição espacial das atividades de educação faz com que o mapeamento da acessibilidade seja diferente: ele é maior na região oeste da cidade, onde há uma maior concentração de escolas públicas. Como também foi observado nos mapas para empregos, a acessibilidade é maior próxima aos corredores de ônibus.

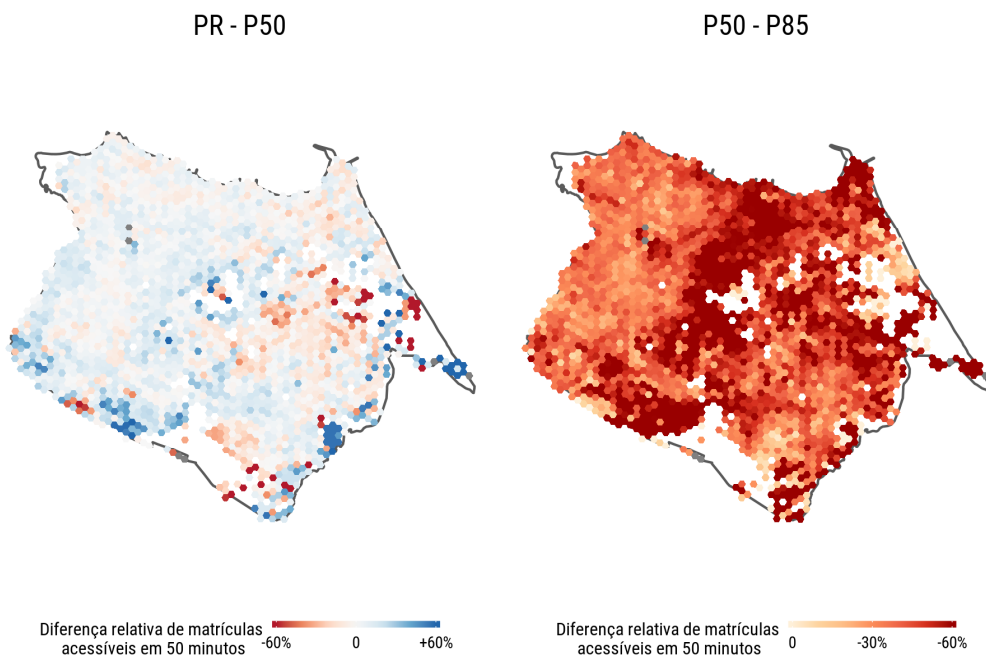
A análise dos *boxplots* mostra uma realidade diferente da encontrada da análise de acesso a empregos. A mediana agora é maior para o P50, com cerca de 78 mil matrículas acessíveis em comparação com 75 mil para o PR - o que representa uma diferença de 4% entre os cenários, menor do que para trabalho. Isso mostra como o acesso às oportunidades de educação é menos impactado pela variação do nível de serviço que acontece na realidade operacional da rede de transporte público, o que ocorre em larga medida porque a distribuição espacial das escolas é menos concentrada espacialmente do que a distribuição de empregos.

Da mesma forma que na seção anterior, a Figura 6 mostra as diferenças PR - P50 e P50 - P85, agora considerando o indicador de oportunidades cumulativas para educação. O retrato mostrado no mapa PR - P50, à esquerda, não permite identificar claramente um padrão espacial de variabilidade na acessibilidade, sem sinais destacados de diferenças entre os cenários. Isso pode significar que o acesso às zonas com maior concentração de matrículas está menos sujeito a congestionamentos e atrasos dos veículos, levando a uma maior pontualidade das viagens, o que resulta em diferenças pouco significativas na estimativa do indicador de acessibilidade.



**Figura 5.** Distribuição espacial (acima) e não-espacial (abaixo) dos indicadores de acessibilidade para educação para cada um dos GTFS

O mapa da diferença P50 - P85 para oportunidades de educação, por sua vez, apresenta um padrão semelhante ao encontrado no mapa P50 - P85 para trabalho: as zonas localizadas nas bordas da área de concentração de atividades (que nesse caso, são matrículas) são as que mais sofrem com a dispersão do nível de serviço da rede. Os números mostram, no entanto, que o P50 superestima a acessibilidade em mais de 60% para aglomerações localizadas no entorno das áreas concentradoras de matrículas, majoritariamente no lado oeste da cidade. Números de diferença absoluta entre os cenários mostram uma superestimação mediana de 27 mil matrículas (9,6% das matrículas totais da cidade), com um 75º percentil de 42 mil (15% do total).



**Figura 6.** Diferença relativa de matrículas acessíveis para comparação GTFS Programado x GTFS Empírico P50 (à esquerda) e GTFS Empírico P50 x GTFS Empírico P85 (à direita)

### 4.3. Impactos da variabilidade na avaliação de intervenções de transportes

A partir dos resultados discutidos acima, pode-se questionar como a utilização de dados observados de tempos de viagem da frota poderia impactar a tomada de decisões em intervenções no sistema de transportes, apoiada em medidas de desempenho da acessibilidade às atividades urbanas. Como observado para atividades de trabalho, dados de GTFS Programado tendem a superestimar a acessibilidade em zonas de entrada do centro da cidade em até 100 mil empregos. Intervenções e análises a partir desses dados tenderiam a superestimar esses valores, podendo levar a decisões de intervenções equivocadas. Um exemplo que pode ser dado é sobre a decisão de intervir construindo um corredor misto de transporte em vez de um corredor exclusivo de ônibus. Num contexto de horários programados, uma análise *ex-post* poderia concluir que houve ganhos de acessibilidade bem semelhantes entre as alternativas, visto que os horários programados tendem a não incorporar efeitos de *bus bunching* e consequentes atrasos. Quando horários observados são incorporados, a diferença entre um cenário de tráfego misto e outro de vias exclusivas de ônibus tende a ser incorporada de forma mais precisa.

Na questão da comparação P50 x P85, é importante observar que a utilização dos valores centrais de tempo de viagem pode superestimar consideravelmente a acessibilidade quando não se leva em conta a dispersão dos mesmos. Para uma situação de proposição de intervenções, a análise da temática fornece informações sobre onde as intervenções podem ser priorizadas: certas áreas estão sujeitas a grande variação na sua acessibilidade, então o aumento da capacidade e/ou a construção de infraestrutura exclusiva para transporte público terá um grande impacto. Na metodologia de avaliação de alternativas, defende-se que utilizar os valores de tempo de viagem P85 oferece uma maior segurança tanto em avaliações *ex-ante* como em avaliações *ex-post*. Portanto, a consideração da dispersão da medida de desempenho no esforço de avaliação de alternativas possibilita uma análise de acessibilidade mais robusta e condizente com as situações de variabilidade dia-a-dia na oferta de transporte público enfrentada pelos usuários.

## 5. CONCLUSÕES

Este estudo teve como objetivo geral analisar se estimativas de acessibilidade por transporte público para trabalho e educação divergem entre si quando calculadas com base em dados programados de GTFS e em dados corrigidos por GPS, incorporando a dispersão do nível de serviço observada entre dias na rede. Para isso, foi proposto um método de agregação de tempos de viagem a partir de um grande período de dados arquivados de GPS para reconstruir arquivos GTFS considerando a variabilidade do nível de serviço do sistema.

O método desenvolvido permite reconstruir as viagens agendadas de transporte público em GTFS com informações reais de velocidade dos veículos e tempos de viagem a partir de dados de GPS. O método foi aplicado a 19 dias de dados de GPS de Fortaleza. Os dados de GTFS da cidade foram reconstruídos considerando-se a mediana do tempo de viagem entre cada trecho de parada (P50), prática padrão adotada em estudos anteriores, e considerando-se o 85º percentil (P85), que permite captar com mais robustez a dispersão da performance do sistema de transporte no seu dia-a-dia. Os dados de GTFS Programado e os GTFS P50 e P85 foram então usados para calcular indicadores cumulativos de acessibilidade para atividades de trabalho e educação. Os resultados mostram que dados de GTFS programado superestimam a performance do sistema de transporte, podendo superestimar os níveis de acessibilidade a empregos e escolas em mais de 50% em alguns casos quando comparados aos dados de GTFS Empírico P50. Mesmo o uso de GTFS Empírico P50, que é o defendido na literatura recente, tende

a superestimar os níveis de acessibilidade quando comparados aos dados de GTFS Empírico P85, que incorporam a variabilidade diária do nível de serviço da rede.

Uma limitação a ser ressaltada na aplicação para Fortaleza decorre de não haver uma correspondência perfeita entre os dados de GPS e GTFS para diversas linhas e veículos. Isso ocasionou um descarte de dados de localização principalmente para linhas que servem a bairros periféricos, geralmente de baixa renda e de baixa densidade populacional. Caso os dados de GPS dessas linhas tivessem sido incorporados os resultados provavelmente amenizariam as diferenças entre as estimativas de acessibilidade com GTFS e GPS. Outra limitação importante é a não incorporação nesse estudo das linhas de média/alta capacidade (Linhas Sul e Oeste do metrô e VLT Parangaba-Mucuripe), em virtude da indisponibilidade desses dados.

Recomenda-se que estudos futuros que utilizem dados de GTFS para estimar indicadores de acessibilidade em avaliações *ex-ante* e *ex-post* de intervenções de transportes busquem incorporar dados observados de tempos de viagem a partir do rastreamento da frota, não só para melhorar a qualidade da estimação, mas principalmente para permitir uma análise de sensibilidade da tomada de decisão à sua variabilidade. Senão, que pelo menos reconheçam as limitações que a utilização de um GTFS programado impõe.

## REFERÊNCIAS

- Arbex, R. O.; B. B. Alves e M. A. Giannotti (2016) Comparing Accessibility in Urban Slums Using Smart Card and Bus GPS Data. *TRB 2016 Annual Meeting*, Washington, DC, Estados Unidos.
- Arbex, R. O. e C. B. da Cunha (2016) 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. *Transportes*, v. 24, n. 4, p. 21. DOI: 10.14295/transportes.v24i4.1008
- Farber, S. e L. Fu (2017) Dynamic public transit accessibility using travel time cubes: Comparing the effects of infrastructure (dis)investments over time. *Computers, Environment and Urban Systems*, v. 62, p. 30–40. DOI: 10.1016/j.compenvurb-sys.2016.10.005
- Geurs, K. T. e B. van Wee (2004) Accessibility evaluation of land-use and transport strategies: Review and research directions. *Journal of Transport Geography*, v. 12, n. 2, p. 127–140. DOI: 10.1016/j.jtrangeo.2003.10.005
- Hansen, W. G (1959) How Accessibility Shapes Land Use. *Journal of the American Planning Association*, v. 25, n. 2, p. 73–76. DOI: 10.1080/01944365908978307
- IBGE (2019) *Estimativas da População Residente no Brasil e Unidades da Federação*. Instituto Brasileiro de Geografia e Estatística.
- Lei, T. L. e R. L. Church (2010) Mapping transit-based access: Integrating GIS, routes and schedules. *International Journal of Geographical Information Science*, v. 24, n. 2, p. 283–304. DOI: 10.1080/13658810902835404
- Mandelzys, M. e B. Hellenga (2010) Identifying Causes of Performance Issues in Bus Schedule Adherence with Automatic Vehicle Location and Passenger Count Data. *Transportation Research Record: Journal of the Transportation Research Board*, v. 2143, n. 1, p. 9–15. DOI: 10.3141/2143-02
- Mavoa, S.; K. Witten; T. McCreanor e D. O'Sullivan (2012) GIS based destination accessibility via public transit and walking in Auckland, New Zealand. *Journal of Transport Geography*, v. 20, n. 1, p. 15–22. DOI: 10.1016/j.jtrangeo.2011.10.001
- Mayaud, J. R.; M. Tran; R. H. Pereira e R. Nuttall (2018) Future access to essential services in a growing smart city: The case of Surrey, British Columbia. *Computers, Environment and Urban Systems*, v. 73, p. 1–15. DOI: 10.1016/j.compenvurb-sys.2018.07.005
- Mazloumi, E.; G. Currie e G. Rose (2010) Using GPS Data to Gain Insight into Public Transport Travel Time Variability. *Journal of Transportation Engineering*, v. 136, n. 7, p. 623–631. DOI: 10.1061/(ASCE)TE.1943-5436.0000126
- Owen, A. e D. M. Levinson (2015) Modeling the commute mode share of transit using continuous accessibility to jobs. *Transportation Research Part A: Policy and Practice*, v. 74, p. 110–122. DOI: 10.1016/j.tra.2015.02.002
- Pereira, R. H. (2019) Future accessibility impacts of transport policy scenarios: Equity and sensitivity to travel time thresholds for Bus Rapid Transit expansion in Rio de Janeiro. *Journal of Transport Geography*, v. 74, n. October, p. 321–332. DOI: 10.1016/j.jtrangeo.2018.12.005
- Pereira, R. H. M.; C. K. V. Braga; B. Serra e V. Nadalin (2019). Desigualdades socioespaciais de acesso a oportunidades nas cidades brasileiras, 2019. *Texto para Discussão Ipea*, 2535. Instituto de Pesquisa Econômica Aplicada (Ipea). Disponível em <http://repositorio.ipea.gov.br/handle/11058/9586>
- Stepniak, M.; J. P. Pritchard; K. T. Geurs e S. Goliszek (2019) The impact of temporal resolution on public transport accessibility measurement: Review and case study in Poland. *Journal of Transport Geography*, v. 75, n. January, p. 8–24. DOI: 10.1016/j.jtrangeo.2019.01.007

- Tornqvist, L.; P. Vartia e Y. O. Vartia (1985) How Should Relative Changes Be Measured? *The American Statistician*, v. 39, n. 1, p. 43-46. DOI: 10.2307/2683905
- Wessel, N.; J. Allen e S. Farber (2017) Constructing a routable retrospective transit timetable from a real-time vehicle location feed and GTFS. *Journal of Transport Geography*, v. 62, n. January, p. 92-97. DOI: 10.1016/j.jtrangeo.2017.04.012
- Wessel, N. e S. Farber (2019) On the accuracy of schedule-based GTFS for measuring accessibility. *Journal of Transport and Land Use*, v. 12, p. 475-500. DOI: 10.5198/jtlu.2019.1502