




A weighted travel time index based on data from Uber Movement

Renato S. Vieira^{1,2*}  and Eduardo A. Haddad¹

*Correspondence:

renato.sv.1988@gmail.com

¹NEREUS, University of São Paulo, Av. Prof. Luciano Gualberto, 908, FEA I, 05508-900, São Paulo, Brazil

²Catholic University of Brasília, SGAN 916 Módulo B Avenida W5—Asa Norte, 70790-160, Brasília, Brazil

Abstract

In this paper, we combine data from Uber Movement and from a representative household travel survey to construct a weighted travel time index for the Metropolitan Region of São Paulo. The index is calculated based on the average travel time of Uber trips taken between each pair of traffic zone and in each hour between January 1st, 2016 to December 31, 2018. The index is weighted based on trips reported in a household travel survey that was designed to be statistically representative of all trips made in the city during a typical business day. We show that the index has a strong correlation with traditional measures of congestion, however, with a broader coverage of the road network. Finally, we use the index to run a multivariate ex-post analysis that estimates the effect of different events on traffic congestion in the city, including holidays, public transit strikes, road shutdowns, rain and major sport events.

Keywords: Traffic congestion; Travel time index; E-hailing data

1 Introduction

Traffic congestion is a major component of transport systems efficiency. With higher levels of congestion, more time is spent in traffic and less time is available for productive activities and for leisure [1]. Therefore, congestion represents a major source of costs and inefficiency for urban economies. Because of that, measuring congestion is a key element for monitoring and evaluating transport systems and for supporting policy decisions aimed to improve them. Traditional methods for measuring congestion such as loop detectors are important tools for road segment monitoring and for traffic signal control. However their coverage is limited to the locations where they are placed, thus, these are not ideal tools for tracking congestion with a highly granular temporospatial coverage [2]. Measuring congestion at the personal trip level requires other types of tools, such as probe vehicles or taxis. However, such methods are costly and are not commonly available in developing country cities where traffic agencies face stricter financial constraints [3].

With technological improvements and the ubiquitous spread of cellphones with geographical tracking and transportation applications, other types of in vehicle data are now broadly available, creating the opportunity for the development of alternative methods to measure characteristics of transport systems with detailed coverage both in terms of time

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

and space and with costs that represent a fraction of those required to install and maintain traditional monitoring structures. One example of this new type of data is the information compiled by the Uber Movement website,^a which provides the average travel time of trips made by vehicles from Uber, a leading e-hailing company. This dataset is available for several major cities throughout the world, and it includes average travel time of trips by hour and by pair of neighborhood origin and destination.

In this paper, we explore data from the Uber Movement (UM) website and combine it with a representative household travel survey for the Metropolitan Region of São Paulo^b (MRSP) to create a virtually cost-free traditional^c Travel Time Index (TTI) that estimates trip delays due to congestion experienced by residents of São Paulo in every hour throughout the last three years and in almost all neighborhoods of the city. We compare this index with a traditional congestion measure calculated by the government of São Paulo, and we show that while there is a strong correlation between the measurements, our index covers a broader set of roads and is more easily translated to actual travel time losses. Finally, we use the TTI to estimate a multivariate model that evaluates the association between different types of events with congestion, identifying the direction and magnitude of those associations.

The remaining of this paper is divided as follows: Sect. 2 describes the UM data and the household travel survey used to construct the index. On Sect. 3 we explain the index and present some of its descriptive characteristics. Section 4 shows the multivariate analyses and Sect. 5 concludes.

2 Data

2.1 Uber Movement

The UM project provides data about travel times of trips made by Uber vehicles in selected cities throughout the world. Cities are divided into neighborhoods according to official traffic zones or administrative boundaries, and the dataset available in the website includes the average travel time of Uber trips made between city neighborhoods at a given time interval.^d That is, for a neighborhood of origin o , and a neighborhood of destination d , if there were a minimum number of Uber trips between those neighborhoods during time period p , then the website includes the average travel time t_{odp} of these trips.

In the case of São Paulo, the metropolitan area is divided into the 517 traffic zones (TZs) defined by the 2017 Origin-Destination Household Travel Survey (OD17) carried by the city's Subway Company (described in further details later in this Section), and average travel times are available from January 1st, 2016 to the most recent finished quarter. While different levels of data aggregation can be extracted directly from the UM website, we requested from the project's maintainers all the hourly data, for all neighborhood pairs of the MRSP in all dates between January 1st 2016 to December 31, 2018. This dataset includes the average travel time observed in 370 million combinations of origin-destination-date-hour. It covers almost all dates^e and all hours from January 1st 2016 to December 31, 2018 (1094 dates and 26,256 hours) and includes 98,063 unique TZ pairs.

2.2 The 2017 São Paulo household travel survey

The UM data is restricted to travel times. However, our goal is to create a congestion measure that accounts for the travel patterns of residents. For example, in the early hours of the day, there is a larger flow of travelers going from residential neighborhoods to the central business district. Therefore, higher levels of congestion in those routes in the morning

affect more individuals than an equivalent level of congestion in those same routes during the evening.

To account for these differences in flows, we use data from the 2017 Origin Destination Household Travel Survey (OD17) carried by the city's Subway Company. The survey interviewed 31,487 households in the MRSP between June 2017 to October 2018, and collected information about 157,992 trips made by 86,318 individuals in the business day immediately before each interview [4]. The survey was designed to be statistically representative for the population of the whole metropolitan region, so each observation included in the survey is associated with a survey weight that can be used to extrapolate results to the overall population of São Paulo.

Besides sample weights, the information from the survey used in this study includes: the TZs of origin and destination of each representative trip (the TZs from the survey are the same as the TZs used by Uber Movement), the transport mode that was used, the departure time. The survey also includes information about travelers' demographics and trip motivation [4].

3 The weighted travel time index

A congestion measure has the objective of allowing quantitative assessments for transportation planners and information to the general public and policy makers. Desirable characteristics of these measurements include: easiness of communication, applicability to different geographical scales, comparability to a certain standard, be measured in a continuous scale, be based on travel time data and be able to describe very congested conditions [5]. One type of metric that satisfies all these conditions and that has been commonly used both by academic works and other types of publications^f are the travel time indexes (TTIs), that are calculated by comparing the travel time of a given trip relative to the expected duration of that same trip under a certain baseline, usually the free-flow conditions. Formally, the index is calculated by the formula:

$$tti_{odp} = \frac{t_{odp}}{t_{od}^*} - 1, \quad (1)$$

where the term tti_{odp} represents the TTI of trips made between a point of origin o and a point of destination d during period p . On the right side of the equation, the term t_{odp} represents the travel time observed for this route in period p , while t_{od}^* indicates the travel time of that same route in free flow conditions. For example, suppose that a certain trip takes on average 32 minutes during the morning peak on business days, and that this same trip takes 20 minutes in free-flowing conditions. In this case, the TTI would be calculated as $32/20 - 1 = 0.6$. That is, the TTI indicates that during the morning peak, this route has an average level of congestion of 60%.

A key advantage of this metric is that it has a simple and intuitive interpretation. Another important characteristic is that the indicator can be aggregated both in time and in space, allowing the analysis and comparison of different regions and different periods. In addition, the index aggregation may be weighted to consider the different travel volumes between regions at different points in time. The aggregation of the TTI can be described by the following formula:

$$TTI_{RP} = \sum_{odp} \left(\frac{t_{odp}}{t_{od}^*} \times \frac{V_{odp}}{V_{RP}} \right) \mid (o, d \in R) \& (p \in P), \quad (2)$$

where TTI_{RP} is the aggregated TTI for region R over a time period P . Besides that, v_{odp} is the number of trips observed between the traffic zones o and d that make up region R , and all periods p that compose P . V_{RP} is the total number of trips between all traffic zone pairs within region R and all the periods that make up the aggregated period P . For example, if o and d are traffic zones, and p are hours of the day, P can be defined as a certain date, and R as a Metropolitan Region as a whole. The result of this aggregation can be directly interpreted as the average level of congestion among all trips made in region R throughout the whole period P .

3.1 Adjustments for calculating the weighted TTI using UM data and the OD17

From the elements of Equation 2, UM data includes the average travel time by hour for each pair of TZ in the SPMR (t_{odp}). So, to calculate the weighted TTI, we still need: 1) the free-flow time between each pair of TZ (t_{od}^*); 2) the number of trips made between each TZ pair at each period (v_{odp}) and the corresponding aggregation (V_{RP}). Next, we describe how each of these elements are estimated in our calculation of the index.

3.1.1 Free-flow travel time (t_{od}^*)

The free-flow travel time is the time that a trip would take if roads were completely free of other vehicles and other factors causing slowness, thus being a theoretical measure. While simulation methods could be used to estimate this metric for the TZ pairs included in our study, such models would require detailed information about the road network such as speed limits and the distribution of origins and destination exact coordinates of trips, both of which are not easily available. Therefore, to overcome these limitations, we estimate free-flow travel time using the own UM data and the assumption that there are periods of the day when observed travel times approximate free-flow conditions, most often during late night hours.

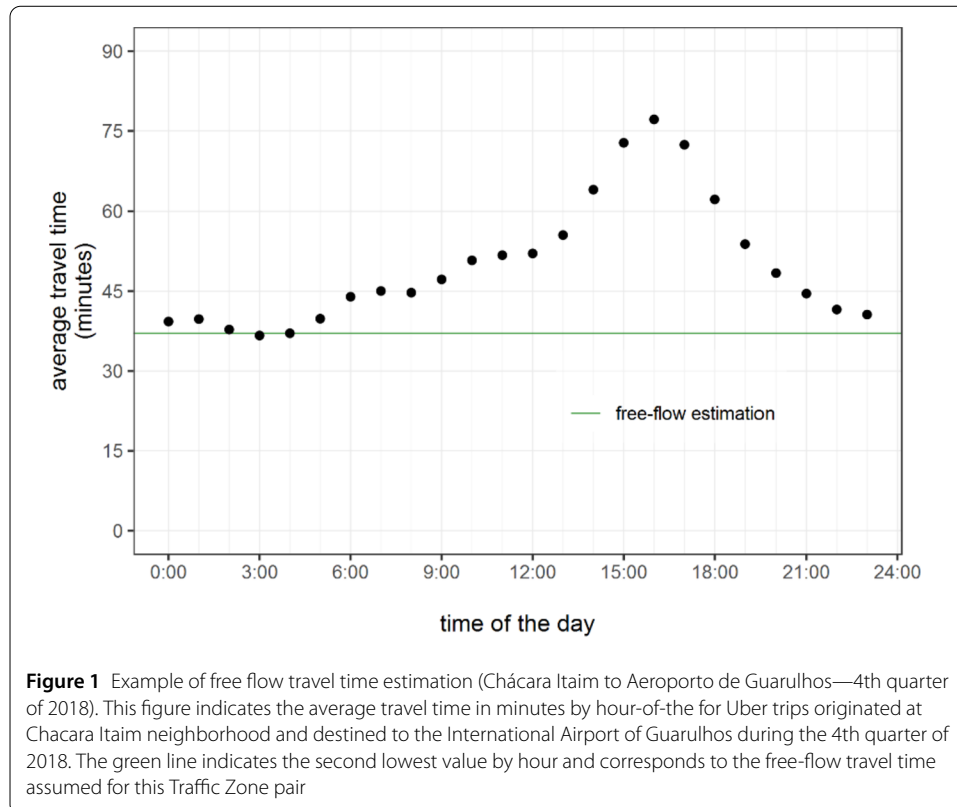
Based on these assumptions, we define free-flow travel time for each TZ pair as the second lowest average travel time value per hour during the most recent quarter of data from UM. The second lowest value is used instead of the first lowest to avoid selecting eventual outlier observations from pairs with fewer Uber trips.⁸ The most recent quarter is used because as Uber expands its services, the density of data from each TZ pair also increases.

To illustrate this approach, Fig. 1 shows the average travel time by hour in a selected TZ pair in the MRSP during the last quarter of 2018. As expected, average travel times by hour are lower during late night hours and increase during peak periods. Given the procedure described above, the average travel time at 4 am is selected as the free-flow approximation for this TZ pair.

3.1.2 Number of trips (v_{odp})

In order to identify the travel flows between each TZ during each period, we use the micro-data from the OD17 Survey. Although the survey is designed to be statistically representative for the MRSP, there are two important practical constraints that need to be considered in the use of this data to calculate the weighting of the TTI:

- 1 The number of TZ pairs observed on UM is not constant; there are days and times when travel time information between a given TZ pair does not exist on the platform because the number of trips made with the Uber application is not enough for the inclusion of the average travel time in the dataset.



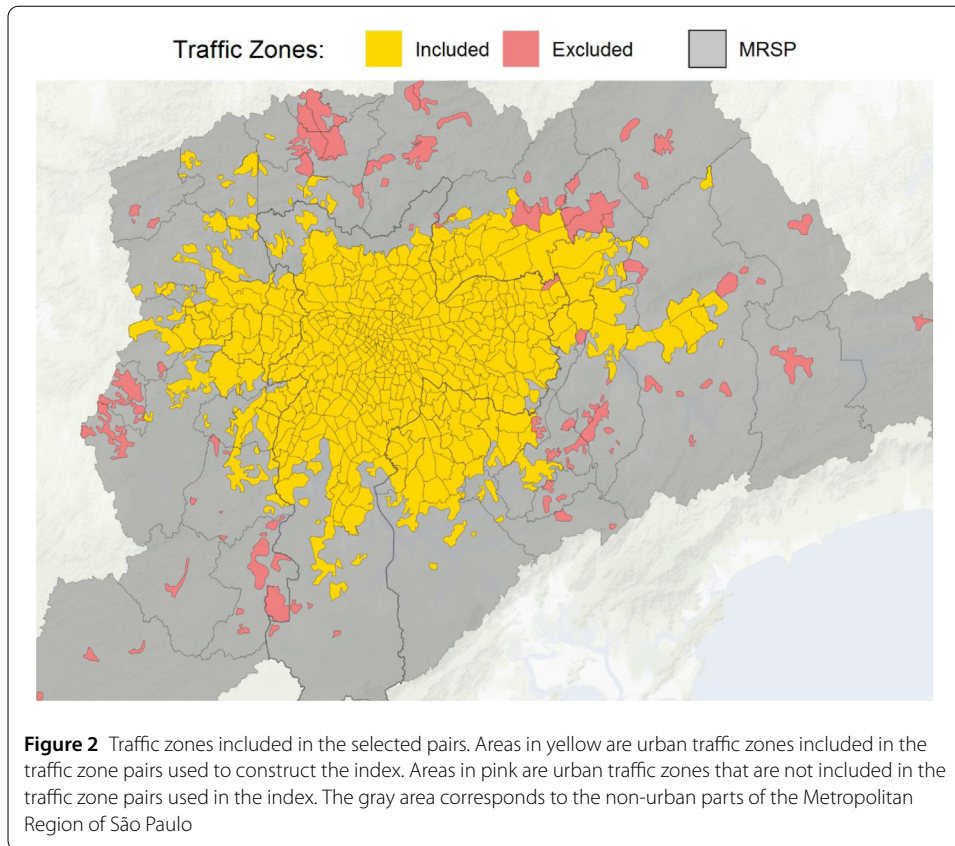
2 The OD17 Survey is not statistically representative at the level of TZ pairs per hour.

The Survey includes information about 157,992 trips made in a typical working day, however, there is a total of 6.4 million combinations of TZ pairs \times hours (517 TZs \times 517 TZs \times 24 hours), therefore the Survey data is not sufficiently dense to inform about travel flows in the same disaggregation level as the travel times from UM.

Regarding constraint 1, the main issue is that, with the expansion of Uber's activities, a direct comparison of congestion levels between distinct periods is not necessarily valid. If the TTI is calculated without considering the differences in composition over time, the results of the analysis may be biased. For example, it is possible that the pairs observed in 2016 are mostly in the more central regions of the MRSP where congestion levels are naturally higher. On the other hand, it is possible that in the case of the observations from 2018, the proportion of suburban TZ pairs become larger. In this case, even if the congestion levels were unchanged in the MRSP, a TTI that does not consider the composition difference between the periods would indicate a congestion drop over time.

To work around this problem, the solution is to keep constant the TZ pairs included in the index construction. We selected the pairs observed in at least three quarters of the days in the analyzed period. With this criterion, 23,807 pairs were selected, and they represent 28.8% of motorized trips in the MRSP according to according to the OD17. Figure 2 shows that although the selected pairs correspond to only 8.9% of the total possible combinations between pairs of zones, they cover practically the entire urbanized area of the MRSP, except for some of the most isolated districts.

As for the second constraint, that is, the non-statistical representativity of the OD17 Survey at the level of TZ pairs by hour, the workaround was to approximate the weighting

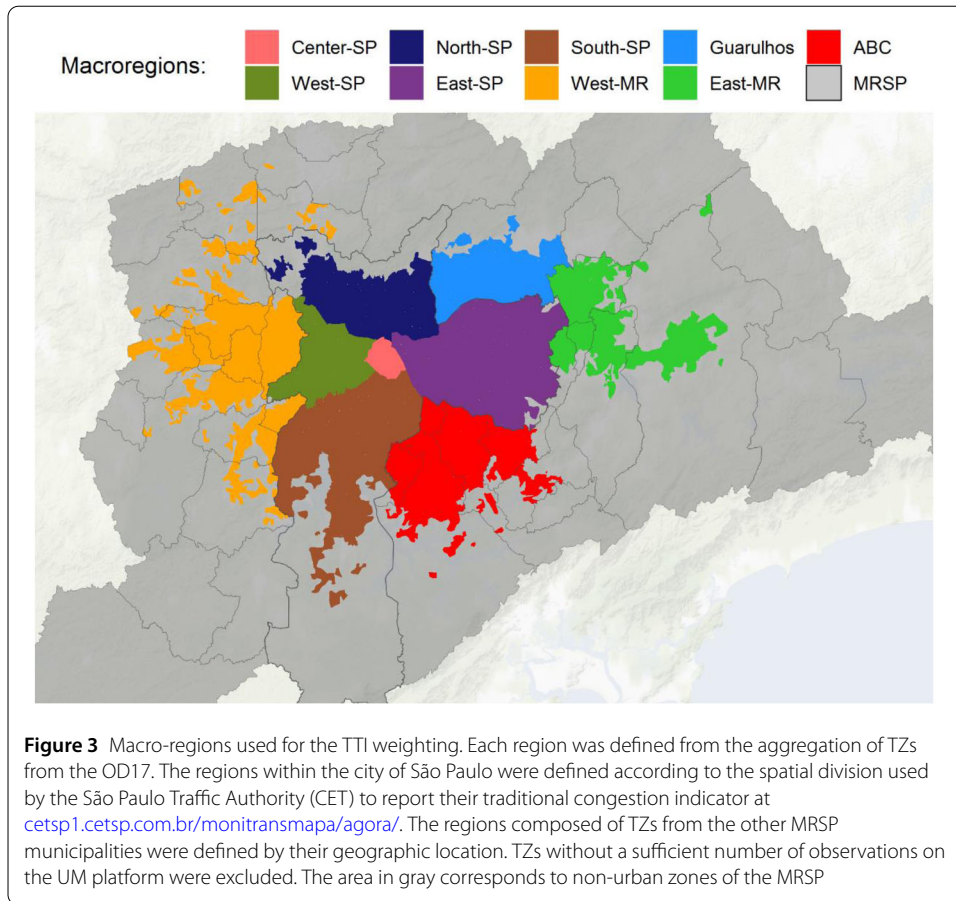


calculation using a higher level of temporospatial aggregation. Specifically, the MRSP was divided into nine macro-regions and the weights were calculated based on the number of trips observed in the OD17 between pairs of macro-regions during different periods of the day^h rather than the number of trips between each pair of TZ by hour, thus ensuring the statistical representativity of the patterns observed between each pair used for the weighting. The macro-regions used in the study were defined according to Fig. 3.

Finally, to ensure that the proportion of trips between each macro-region is constant even if the number of TZ pairs change over time in the UM data, an adjustment factor was added to the TTI formula in order to maintain constant the total weight of each macro-region pair.

$$w_{odp} = \frac{v_{\bar{O}\bar{D}p} / N_{\bar{O}\bar{D}p}}{V_{RP}}, \tag{3}$$

where $v_{\bar{O}\bar{D}p}$ is the total number of trips observed in the OD17 Survey between the OD macro-region pair during period P , such as $o \in O$ and $d \in D$. In addition, $N_{\bar{O}\bar{D}p}$ is the total number of TZ pairs that make up the macro-region $\bar{O}\bar{D}$ and which have travel time information in period p on the UM platform. This adjustment ensures that even if the number of TZ pairs fluctuate throughout time, the TTI results will always correspond to the static aggregated traffic volumes observed in the OD17. That is, if a pair of macro-regions contains 10% of the total travel flows in the OD17, then the sum of the weights of the TZ that make up that macro-region pair will always be equal to 10%, and this equivalence will be independent of the number of TZ pairs observed in the UM data in a given day. Therefore,



the final adjusted calculation of the weighted TTI is described by the following formula:

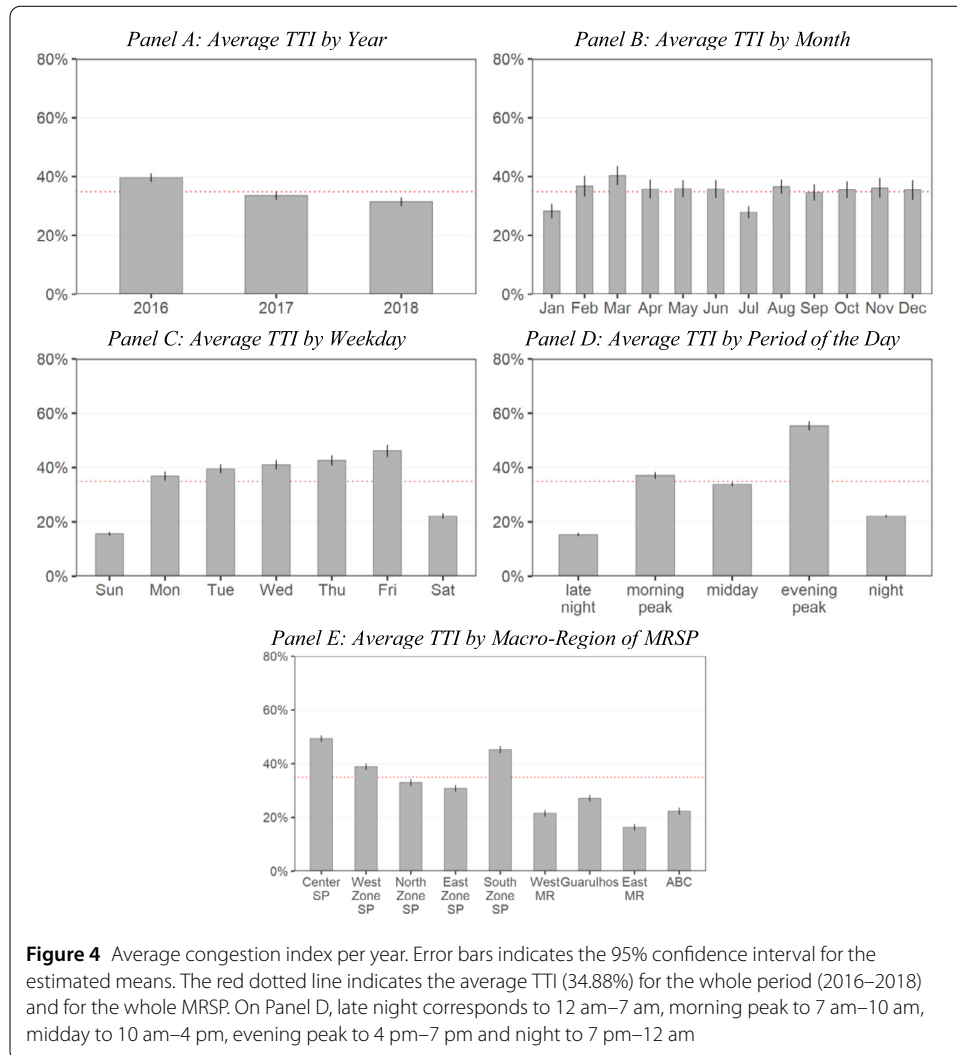
$$TTI_{RP} = \sum_{odp} \left(\frac{t_{odp}}{t_{od}^*} \times w_{odp} \right). \tag{4}$$

3.2 TTI descriptive results

Based on the formula described in the previous sub-section, the TTI was calculated for all dates from January 1st, 2016 to December 31, 2018. For each day, the index was also calculated for each period of the day and for each macro-region of the MRSP.

The average value for the total TTI was of 34.88%, indicating that on average, trips made in São Paulo are 34.88% longer than if made on free-flow conditions.¹ Fig. 4 plot the average TTI by different types of segmentation. In all plots, the dotted red line indicates the overall mean of 34.88. Panel A shows the average TTI by year (2016, 2017 and 2018), showing a decrease of approximately 6.1 percentage points between 2016 (39.6%) to 2017 (33.5%), and a new reduction of 2.1 points in 2018 (31.4%). Panel B shows the within year seasonality of the TTI mean by comparing the average levels of the index by month, showing that in the months of school holidays (January and July) congestion levels are below average (28.3% and 27.9%, respectively). Meanwhile, March is the month with the highest average TTI (40.3%).

On Panel C, we show the dynamics of congestion within the week. On weekends, the mean TTI is well below average (15.6% on Sundays and 22.2% on Saturdays). On week-



days, there seems to be an increasing pattern, with Friday being the most congested day (46.2%). Panel D shows the dynamics of congestion within the different periods of the day. The afternoon peak, defined as the period between 4–7 pm, has the highest average TTI (55.4%). Meanwhile, the night and late-night periods have lower average value, with mean TTIs of respectively 22.2% and 15.3%.

Finally, Panel E shows the average TTI within each of the MRSP macro-region. This comparison shows that the South, West and Central regions of São Paulo have the highest congestion averages (45.5%, 37.4% and 44.2%, respectively). On the other hand, the regions outside the capital (West of MR, Guarulhos, East of MR and ABC) present congestion levels below the overall average (21.9%, 26.4%, 16.3% and 24.8%).

3.3 Total time spent due to congestion

Based on the TTI and the OD17 data, it is possible to estimate with a simple back-of-the-envelope calculation the average time lost due to congestion in a typical weekday in São Paulo. As shown in the previous section, the average TTI on weekdays is equal to 41.2%, and according to the OD17, in a typical business day, 3.62 million individuals travel by car in São Paulo spending a total of 5.47 million hours on these trips. Therefore, if we divide

this total number of hours by one plus the average TTI, we have that those same trips would spend only 3.87 million hours if they were always made under free-flow conditions. That is, congestion causes individuals who travel by car in São Paulo to spend, on average, 26.4 more minutes than if all trips were performed under free-flow conditions.

As pointed by Litman (2009), [1], expecting all trips to be performed in free-flow conditions is not something reasonable, especially in dense urban environments. However, the objective of this type of calculation is not to set up a target of potential time savings. Instead, the most important value of this estimation is to translate into easily understandable metrics the potential outcomes of different scenarios. For example, we can calculate that the 6.1 percentage point reduction in the TTI between 2016 to 2017 corresponds to an average reduction of almost 3 minutes per day for all individuals who travel by private cars^j in the MRSP.

3.4 Comparison with traditional congestion measures

To further validate the TTI calculated in this paper, we compare it with a traditional congestion metric that is currently calculated by CET,^k which counts the length of congestion on selected roads by hour. The CET measure is calculated based on the evaluation of technicians in official vehicles or positioned at the top of buildings. Only roads with the largest volume of vehicles are measured, corresponding to just over 800 km of the 17,000 km of the total roadway network in São Paulo^l [6].

Figure 5 shows the scatter plot of our TTI on the y -axis against the CET congestion measure on the x -axis. To make the values comparable, we restricted the TTI to the same

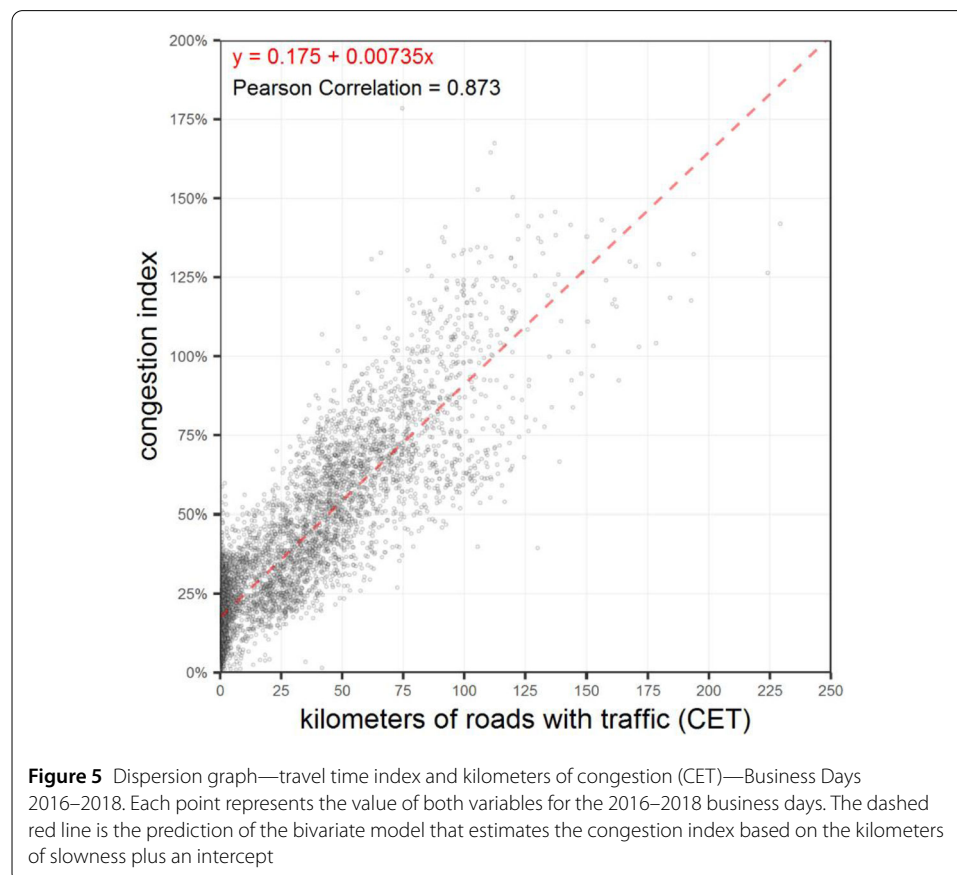
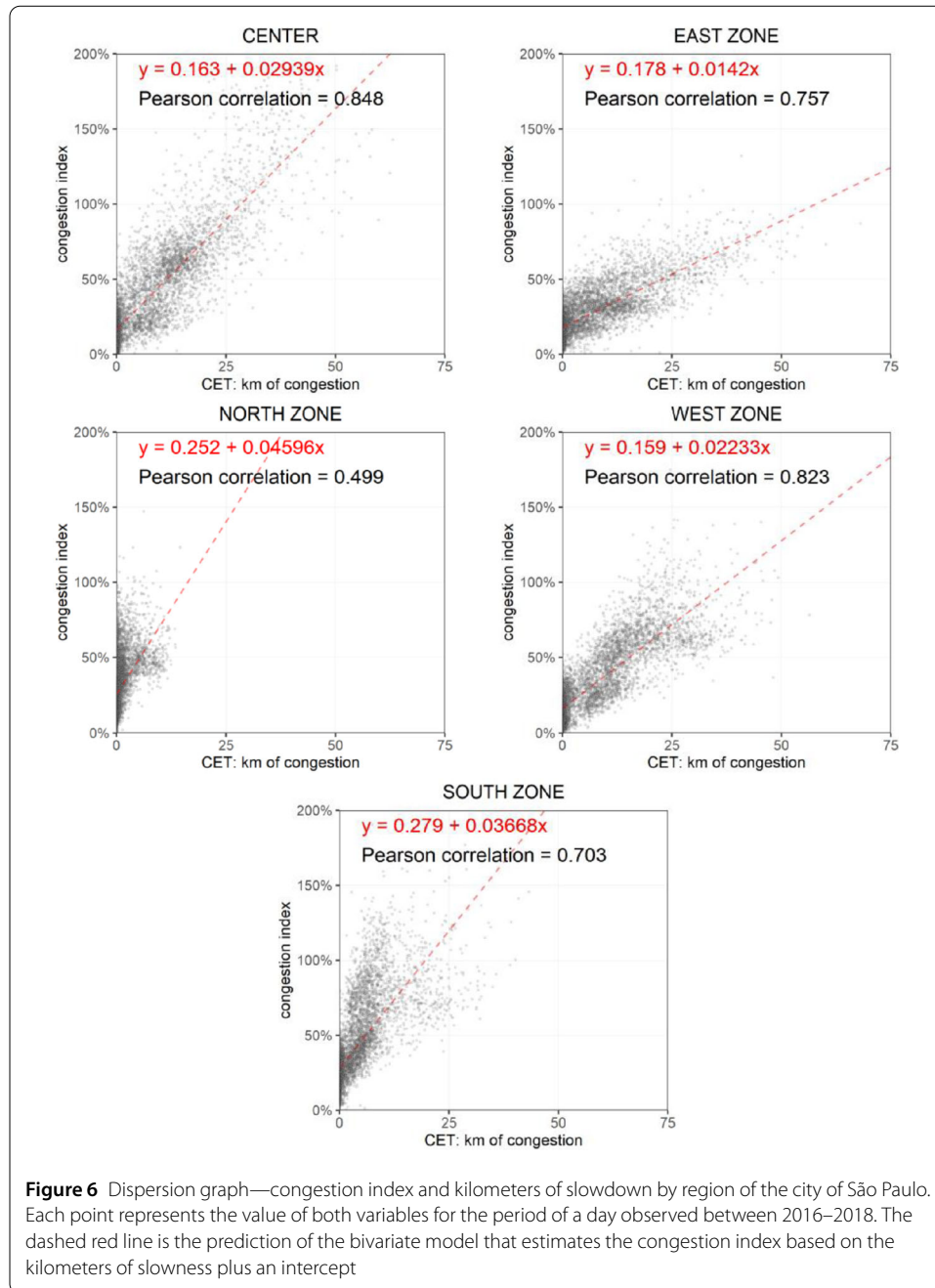


Figure 5 Dispersion graph—travel time index and kilometers of congestion (CET)—Business Days 2016–2018. Each point represents the value of both variables for the 2016–2018 business days. The dashed red line is the prediction of the bivariate model that estimates the congestion index based on the kilometers of slowness plus an intercept



5 regions that are used in the CET measure, and we also restricted the analysis to business days. Therefore, each point in the plot represents the value for both measures by region of São Paulo city and by day. The figure shows a clear and strong positive correlation between the 2 measure (0.873). However, it is worth noticing that the TTI values are less likely to be equal or very close to zero, mainly because the road network is not restricted to certain roads as in the case of the CET measure.

Next, we present on Fig. 6 the separated plots for each region of the city. The results of these analyses show that most series have a high degree of correlation (above 0.7). The only exception is the North Zone, where the correlation coefficient is equal to 0.499, which

can only be considered a moderate positive association between the series. One possible explanation for this difference is the more restricted spatial coverage of the CET indicator in the North Zone.

The high correlation between our TTI and the traditional CET indicator serves as a validation for the measure developed in this study, encouraging its use for analyses that extrapolate the space and time limitations of the CET congestion indicator. While the series are similar, it is important to highlight a key difference between the indicators: the CET measurement is based on the level of traffic flow slowness observed in the main avenues of São Paulo, meanwhile, the TTI is calculated from actual travel times of real trips, therefore it already accounts for drivers routing optimization and adjustments to real traffic conditions. Thus, in addition to having a simpler and more direct interpretation, the TTI results can be directly translated into travel costs.

In the next section, we explore and present additional and more detailed examples of analyses that can be performed using the TTI results that were calculated in this section.

4 Analysing the impact of different events on traffic congestion

There are several factors that can affect the level of congestion in a city like São Paulo, including climatic events, strikes, political demonstrations, festivities, historical trends and other atypical events. Given the granularity of the time series of the congestion index, it is possible to estimate and compare the impact of different events in the city traffic through a multivariate analysis. In the period of our analysis, by way of illustration, we identified the following factors that could potentially affect São Paulo traffic: Rainy days; A national truck drivers' strike between May 21 to May 31, 2018; The 2018 FIFA World Cup (especially when the Brazilian national team was playing); School holidays; The closing of part of Marginal Pinheiros due to the collapse of a bridge (Nov. 15, 2018).

To evaluate the impact of each of these events, we estimate the following multivariate regression model:

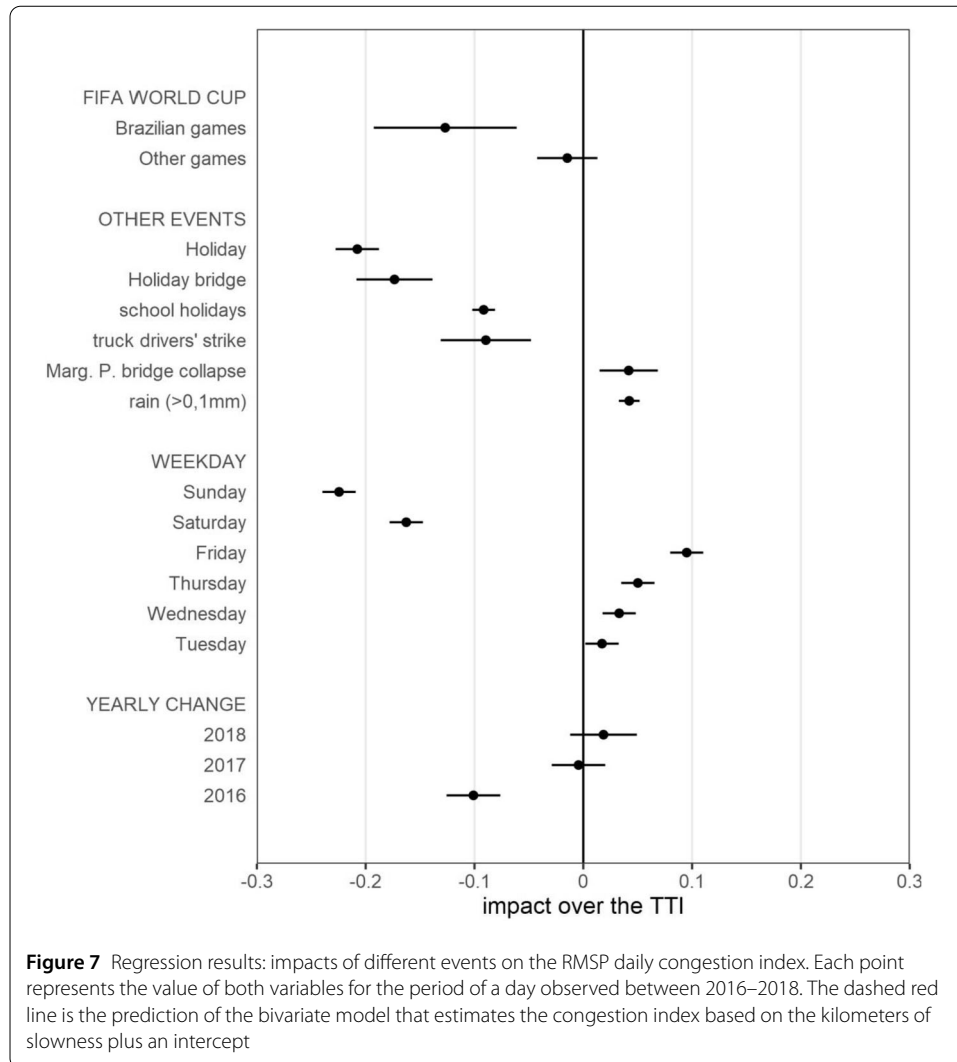
$$C_t = \alpha_t + \delta_t + \beta R_t + \sigma S_t + \gamma_1 W_t + \gamma_2 WB_t + \eta H_t + \theta E_t + \xi SH_t + \omega_t t + \varepsilon_t, \quad (5)$$

where:

- C_t : the congestion index C in the MRSP during each date t ;
- α_t : a vector of year specific intercepts;
- δ_t : a vector of fixed effects associated with each day of the week (Mon, Tue, ...);
- R_t : a dummy variable indicating days with more than 0.1 mm of accumulated rain in the RMSP;
- S_t : a dummy indicating the dates of the truck drivers' strike (May 21–31, 2018);
- W_t : the 2018 FIFA World Cup dates (06/14/2013–07/15/2018)
- WB_t : The Brazilian National team games during the World Cup;
- H_t : holidays;
- E_t : holiday bridges,^m
- SH_t : school holidays (winter and summer);
- ω_t : year specific linear time trends.

The results of this model estimation are presented in Fig. 7:

During the 2018 World Cup, when the Brazilian National was not playing, the average congestion was not much different than usual, averaging 1.4 percentage points below ex-



pected (a nonsignificant result at 5% significance). However, on the days of Brazil’s games, the index was on average 11.2 points below usual.

Regular holidays are associated with a congestion reduction of –20.7 percentage points and holiday bridges have a similar association (–17.3 points). During school holidays, congestion is 9.1 points lower. During the truck drivers’ strike, the circulation of vehicles was greatly reduced due to the lack of fuel at filling stations. So not surprisingly, the congestion index during the strike was 8.9 points lower than usual. In the days following the closing of Marginal Pinheiros due to a bridge collapse, there was an increase of approximately 4.1 points in the TTI. Finally, on days with average rainfall above 0.1 mm per hour, congestion is on average 4.2 points higher than on non-rainy days.

As already noted in Sect. 2, the congestion patterns are heterogeneous throughout the days of the week. The multivariate regression estimated here confirms the same patterns. The reference group are Mondays. Therefore, the results indicate that the other days of the week present an average value of increasing congestion, with the highest value observed on Fridays (9.5 points higher than Mondays). On weekends, the TTI is well below the reference group, respectively –16.2 on Saturdays and –22.4 on Sundays. Finally, the year

specific slopes show that while there was a significant decrease of 10 points in 2016, the TTI remained mostly stationary during 2017 and 2018.

5 Conclusion

As in most large metropolises around the world, traffic congestion is one of the greatest problems faced by the residents of São Paulo, generating economic and welfare losses for residents and for visitors. A first step for addressing the problem of traffic congestion is to measure, monitor and understand the phenomenon. However, one difficulty faced by technicians and researchers, particularly in developing world cities, is the lack of large-scale quantitative measures with high temporospatial granularity.

The TTI built in this study based on data from an e-hailing company aims to provide a new tool for congestion analysis that is virtually free and could potentially be extended to other cities and to other types of analysis. The indicator constructed here differs from traditional measures used in São Paulo by being based on actual travel time from real trips rather than road-based metrics. Because of that, the indicator directly reflects travel time costs while accounting for route adaptation and optimization. Additionally, the index created here suggests a framework for integrating UM data with household travel surveys in order to create a weighting scheme that makes the index results to reflect the travel patterns and average delays experienced by drivers.

Still, the framework proposed in this paper is open to technical improvements and adjustments, such as refinement of free-flow metrics and travel weighing. A natural extension of the framework presented here would be to go beyond a traditional averaged Travel Time Index and to use the integrated data to calculate reliability measures such as a Planning Time Index. We also acknowledge that the multivariate analysis estimated here doesn't fully take advantage of the temporospatial granularity of the index. Therefore, the results presented here represent an initial outline of the possible applications of the index, and other studies using and improving the indicator, as well as exploring the Uber Movement data, are highly recommended. It is hoped that such tools and studies will be used to objectively inform the urban mobility debate and to assist policy makers in their decisions.

Acknowledgements

We are grateful for helpful comments from Daniel A. Brent. We thank participants at the NEREUS-USP Seminars, the North American Regional Science Association Meeting (NARSC, 2019), and the International Workshop "CITIES4PEOPLE: Towards Smart, Safe and Sound Cities (JADS, 's-Hertogenbosch, 2019). We thank André Monteiro, Flavia Annenberg, Gabriela Barbosa, Rafael Alloni and Emily Strand from Uber for their technical support and help with the Uber Movement data.

Funding

Eduardo A. Haddad acknowledges financial support from CNPq (Grant 302861/2018-1), and the National Institute of Science and Technology for Climate Change Phase 2 under CNPq Grant 465501/2014-1 and FAPESP Grant 2014/50848-9.

Abbreviations

CET, Companhia de Engenharia e Tráfego, the São Paulo Traffic Authority; MRSP, Metropolitan Region of São Paulo, the 39 municipalities that make up the Greater São Paulo; OD17, A Household Travel Survey carried in 2017 by the São Paulo Subway; TTI, Travel Time Index; TZ, Traffic Zone; UM, Uber Movement, a website that compiles travel time information of trips made by Uber vehicles.

Availability of data and materials

The Uber Movement datasets analysed in this study are available at <https://movement.uber.com/>. The 2017 Household Travel Survey of São Paulo is available at <http://www.metro.sp.gov.br/pesquisa-od/>.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Both authors provided equal contributions to all sections of the paper. All authors read and approved the final manuscript.

Endnotes

- ^a <https://movement.uber.com/>
- ^b With approximately 20.8 million residents, the Metropolitan Region of São Paulo is the largest urban agglomeration of the Southern Hemisphere.
- ^c The Travel Time Index constructed in this paper is based on simple averages of travel time compared to free-flow. The recent literature on congestion measuring tends to place greater emphasis on travel time reliability on top of travel time averages. That is because improvements in reliability improve travelers' value by allowing them to better optimize their allocation of time [7]. Therefore, extending the index created here to account for travel time reliability is a suggested topic for further research.
- ^d Further details about the methodology used to calculate the data is available at the project's technical note available at https://movement.uber.com/_static/76002ded222a46a02ae89f207e91e335.pdf. The release of Uber Movement data is very recent and its caveats and limitations have not yet been formally discussed by independent studies. Wu (2019) [8] compared Uber Movement travel times with predictions from Google Maps API for the Greater Sydney Region, observing that Uber Movement travel times tended to be lower if compared to the Google API predictions. Possible explanations for this difference include differences in travel speed between Uber drivers and other travelers and a conservative time buffer in the case of Google Maps API. An example of academic work using Uber Movement data is Aryandoust (2019) [9], that calculated parking density maps for 39 different cities throughout the world.
- ^e Due to issues in the system of extraction, data from October 27–28, 2018 was not collected.
- ^f Examples of TTIs used or discussed in academic papers include [10–16]. The TTI is also among the performance indicators used in Urban Mobility Report, a periodical that compares congestion in the 494 largest American urban areas [17], and is also the indicator used to construct the TomTom Traffic Index that compares congestion levels in 403 cities across 56 countries [18].
- ^g The second lowest hourly average travel time corresponds to approximately the 8th percentile of the data per pair.
- ^h Late-night (midnight to 7 a.m.), morning peak (7 a.m. to 10 a.m.), midday (10 a.m. to 6 p.m.), afternoon peak (4 p.m. to 7 p.m.) and night (7 p.m. to midnight). These are the periods defined on the global UM platform.
- ⁱ This extrapolation of results to the total population depends on the accuracy of the weights from the household survey. If the weighted distribution of survey participants differs from the actual distribution of the population, then these types of extrapolations not be accurate. However, the household survey used in our study was designed to be representative of the city's population and of trips made by residents in a typical business day, so weights were calculated in order to allow for such extrapolations [4].
- ^j An important limitation of this calculation is that it is restricted to trips made with private cars (47.4% of the trips made by motorized vehicles that circulate on the streets, according to the OD07). Obviously other modes such as buses and bikes are also affected by congestion, but at different scales. Thus, the measure estimated here does not contemplate the total hours lost by all São Paulo citizens, but only by the portion that circulates by car. The estimation of total losses is undoubtedly an interesting exercise, but it would require additional data and analysis, thus being a suggestion for future studies.
- ^k Companhia de Engenharia e Tráfego, the São Paulo Traffic Authority.
- ^l More details about the CET measure can be accessed at <http://cetsp1.cetsp.com.br/monitransmapa/agora/ajuda.htm>.
- ^m Holiday bridges occur on Mondays when Tuesdays are holidays, and Fridays when Thursdays are holidays.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 February 2020 Accepted: 3 August 2020 Published online: 08 August 2020

References

1. Litman T (2009) Transportation cost and benefit analysis. Victoria Transport Policy Institute 31
2. Kalinic M, Krispb JM (2019) Floating car data and fuzzy logic for classifying congestion indexes in the city of Shanghai. Proc Int Cartographic Assoc. <https://doi.org/10.5194/ica-proc-2-57-2019>
3. Vaziri M, Jafarabady R, Bindra SP (2007) Modeling highway congestion index for a developing country: the Iran experience. http://scientiairanica.sharif.edu/article_2988.html
4. METRO (2019) A Mobilidade Urbana da Região Metropolitana de São Paulo em Detalhes São Paulo. http://www.metro.sp.gov.br/pesquisa-od/arquivos/Ebook%20Pesquisa%20OD%202017_final_240719_versao_4.pdf
5. Levinson HS, Lomax TJ (1996) Developing a travel time congestion index. Transp Res Rec 1564(1):1–10
6. CET—Trânsito Agora. <http://cetsp1.cetsp.com.br/monitransmapa/agora/ajuda.htm>
7. Office of Operations (2006) Travel time reliability: making it there on time, all the time Federal Highway Administration, Washington. https://ops.fhwa.dot.gov/publications/tt_reliability/brochure/
8. Wu H (2019) Comparing Google maps and Uber Movement travel time data. Transp Find. <https://doi.org/10.32866/5115>
9. Aryandoust A, van Vliet O, Patt A (2019) City-scale car traffic and parking density maps from Uber Movement travel time data. Sci Data 6(1):1–18

10. Li Z, Hong Y, Zhang Z (2016) Do ride-sharing services affect traffic congestion? An empirical study of uber entry. *SSRN Electron J* 2002:1–29
11. Bertini RL (2006) You are the traffic jam: an examination of congestion measures. In: The 85th annual meeting of transportation research board
12. Mehran B, Nakamura H (2009) Considering travel time reliability and safety for evaluation of congestion relief schemes on expressway segments. *IATSS Res* 33(1):55–70
13. Sweet MN, Chen M (2011) Does regional travel time unreliability influence mode choice? *Transportation* 38(4):625–642
14. Bennecke A, Friedrich B, Friedrich M, Lohmiller J (2011) Time-dependent service quality of network sections. *Proc, Soc Behav Sci* 16:364–373
15. Karim L, Daissaoui A, Boulmakoul A (2017) Robust routing based on urban traffic congestion patterns. In: ANT/SEIT, pp 698–703
16. Chen P, Tong R, Lu G, Wang Y (2018) Exploring travel time distribution and variability patterns using probe vehicle data: case study in Beijing. *J Adv Transp* 2018:3747632
17. Schrank D, Lomax T, Eisele B (2019) 2019 urban mobility report. Texas Transportation Institute [ONLINE]. <https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-report-2019.pdf>
18. TomTom Traffic Index. https://www.tomtom.com/en_gb/traffic-index/

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
