Disparities in travel times between car and transit: Spatiotemporal patterns in cities

Yuan Liao[1,*], Jorge Gil[2], Rafael H. M. Pereira[3], Sonia Yeh[1], and Vilhelm Verendel[4]

[1]Department of Space, Earth and Environment, Division of Physical Resource Theory, Chalmers University of Technology, Gothenburg 41296, Sweden
[2]Department of Architecture and Civil Engineering, Division of Urban Design and Planning, Chalmers University of Technology, Gothenburg 41296, Sweden
[3]Institute for Applied Economic Research (Ipea) - Brazil, Department of Urban, Regional and Environmental studies and policies (DIRUR), Brasilia 70076-900, Brazil
[4]Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg 41296, Sweden
[*]Corresponding author: yuan.liao@chalmers.se

ABSTRACT

Many cities worldwide are pursuing policies to reduce car use and prioritise public transit (PT) as a means to tackle congestion, air pollution, and greenhouse gas emissions. The increase of PT ridership is constrained by many aspects, and among them travel time and the built environment are considered the most critical factors in the choice of travel mode. Our results suggest that using PT takes on average 1.4-2.5 times longer than driving a car. We use real-time traffic data, transit data, and realistic demand estimates to compare travel times by car and PT in four cities (São Paulo, Brazil; Stockholm, Sweden; Sydney, Australia; and Amsterdam, the Netherlands) at high spatial and temporal resolutions. The share of area that favours PT over car use is very small: 0.58% (0.61%), 0.43% (0.46%), 0.98% (1.14%) and 1.12% (1.14%) for the daily average (and during peak hours) for São Paulo, Sydney, Stockholm, and Amsterdam, respectively. The travel time disparity, as quantified by the travel time ratio R (PT travel time divided by the car travel time), varies widely during an average weekday, by location and time of day: there is less disparity near city centres, around PT lines, and during congestion hours. But R becomes extremely large ($R > 5$) at night when few transit services are available. A systematic comparison between these two modes shows that the average travel time disparity is surprisingly similar across cities: $R < 1$ for travel distances less than 3 km, then increases rapidly but quickly stabilises at around 2. Communities and authorities need to be better informed of the detailed spatial and temporal patterns of disparities in travel times between the two modes, as a better understanding of this spatiotemporal variation is crucial to more accurate predictions of residents' mode choices and identify potential areas for improvements in future studies.

Introduction

Increased car use has many negative environmental impacts, including traffic congestion, land-use issues such as parking, increased air pollution, and greenhouse gas (GHG) emissions. Public transit (PT) can provide a low-cost, energy-efficient, less polluting, and socially equitable travel alternative[1,2]. Policymakers worldwide recognise the importance of promoting a mode shift from car to PT in cities as a way to address negative environmental impacts, increase equity[3], and combat climate change[4]. For instance, Berlin is committing 28 billion euros from 2019 onward to improve PT[5]. However, increasing PT ridership and reducing private car use have proven difficult. Despite significant investments, PT ridership has fallen steadily in the United States and elsewhere[6,7]. Growth in PT ridership is constrained by many factors, such as fixed schedules and routes, low population density, and travellers' attitudes[8]. Among these constraints, travel time and the built environment are considered the most critical factors in the choice of transport mode[6,9,10]. A primary driver of PT ridership growth is the reduction of users' perceived marginal cost, including travel time[11–13]. Travel time is also a key performance indicator for the quantification of PT service quality[14]. In a review paper, Redman et al. summarised studies of PT improvement strategies targeting different quality attributes[15] and found that most studies focused on speed as a critical factor for increasing PT ridership. In a New York study, a 15-minute shorter commuting time corresponded to about 25% higher ridership for a rail service[15]. According to a post-trip questionnaire survey of car users, shorter travel time is one of the key factors that would make PT more attractive[16]. While there is strong evidence that car-based travel is often faster than public transit, the spatial and temporal patterns of this time discrepancy is not well-known. This knowledge could help identify opportunities where and when public transit is competitive (time-wise) with automobiles and shed light on the relative transportation disadvantage of members of the community who must depend on public transit.

Robust methods and data for comparing travel times between cars and PT are much needed for real-world assessments to better inform urban planning and policy efforts to encourage transport mode shifts. Despite the growing interest in understanding the disparities in travel times for cars and PT[17], previous studies have largely relied on simplistic assumptions, using constant vehicle speeds[18] and overlooking how travel speeds vary over the course of the day[19]. Recent studies show how transit travel times can vary greatly by route and time of day[20,21], while a growing body of research using GPS and mobile phone data to record travel speed shows that overall transport performance varies significantly by time of day due to the congestion in cities[22–24]. Moreover, most studies compare travel times between cars and PT based on a limited number of predefined departure times[25] and origin-destination pairs for a selected city[26] or region[18], overlooking how both transport demand and transport performance vary by place and time of day. In summary, previous studies generally overlook the different spatial and temporal dynamics underlying car and PT performances, which complicates comparisons of travel time estimates for travel by car and PT[25] and, therefore, constrains the ability of policymakers to make informed decisions.

Information and communication technologies (ICT) and the trend of big data could possibly help to fill these gaps. Rapidly emerging data sources and geographical information systems (GIS) have significantly increased the availability and the amount of data sensed in urban transport systems[27,28], such as traffic speed data, taxi GPS data[28], and PT smart card data[29]. The availability of real-time traffic speed data enables more advanced traveller information systems for route choice and better-informed traffic planning[30]. New data sources, such as HERE Traffic[i] with extensive coverage of cities in 83 countries to date[30], can collect and provide information about real-time road speed, incidents, and accidents. The amount of available data and the level of spatial and temporal details allow more realistic estimates of travel time[24].

Moreover, social media services, e.g. Twitter, are increasingly used to study mobility patterns such as travel demand and travel behaviours[31]. Geotagged social media data (e.g. tweets with GPS coordinates specifying location) have also been found to be a useful proxy for travel demand[32] when compared with actual traffic data[33] and origin-destination data from other data sources, e.g. travel surveys[32,34]. Geotagged tweets have been used to derive the spatiotemporal clusters of urban human activities that are associated with traffic congestion[35].

The rise of open data also supports more realistic travel time calculations. Open data standards such as General Transit Feed Specification (GTFS)[ii] and crowdsourcing initiatives such as OpenStreetMap (OSM)[iii] provide data that can be used to estimate PT travel times with the most up-to-date schedules[36].

The existing literature lacks the integration of real-world data on driving speed and travel demand with the fine characterisation of spatial and temporal patterns across different cities. To bridge this gap in the literature, this study leverages multiple large-scale data sources to capture at a fine resolution the spatiotemporal patterns of how car and PT travel times vary in four different cities: São Paulo, Brazil; Stockholm, Sweden; Sydney, Australia; and Amsterdam, the Netherlands. This study calculates the detailed spatiotemporal variations of travel times for an average weekday to improve the level of resolution at which we can understand the disparity in travel times between PT and car. The study combines multiple data sources, including HERE Traffic data over one year to derive empirical road speed, Twitter data accumulated from the past nine years, up-to-date GTFS transit data, and road networks from OpenStreetMap. Each city is divided into a hexagonal grid system, and travel times are estimated over various times of the day for travel to any highly visited cell (destination), identified as such based on geotagged tweet volumes, from any cell within the system (for more details, see Methods). We calculate travel times by car and by PT in a door-to-door fashion. Within a selected time interval (e.g., 8 am to 9 am), the average travel time of a given origin cell is defined as the mean value of the travel times from that origin to multiple destinations whose frequency of geotagged tweets are used as weights. To quantify the modal disparity of travel time, we use the travel time ratio (R), defined as the travel time by PT divided by the travel time by car for a given origin-destination pair at a certain departure time. Finally, we visualise and analyse the results to demonstrate how car and PT travel times vary spatiotemporally across all the cities studied. We use real-world large-scale datasets to incorporate realistic travel-demand patterns and to provide higher resolution of the spatiotemporal characterisations of private and public transportation systems. In addition, we present a systematic cross-regional comparison of the travel time disparity between car and PT in the four cities studied.

Results

Spatiotemporal variations of travel times

The temporal variation of the citywide average travel times (see Methods) to frequently visited locations is illustrated in Figure 1. The two modes display rather different temporal patterns. For car use, Sydney and Stockholm have two distinct peaks during morning and afternoon rush hours, while São Paulo and Amsterdam show elevated travel time during daytime, but the rush-hour peaks are less pronounced. For PT, the travel time is generally longer from midnight to dawn when services are suspended or

[i] https://www.here.com/
[ii] https://developers.google.com/transit/
[iii] https://www.openstreetmap.org/

infrequent. In São Paulo, close to 80% of the grid is accessible by PT. Stockholm, Sydney, and Amsterdam have less coverage, in succession. Sydney has the most time-variable coverage, measured by the change in the share of grid cells that can access the destinations over the course of a typical weekday. Figure 2 shows the spatiotemporal variation in the travel time ratio (R) for the origin locations in each city. During peak hours, R varies less with location than at other times of day. In Amsterdam, the maximum share of grid cells that can access the destinations is around 40% which is much lower than the other cities (Figure 1). This corresponds to the large "blank" area in Amsterdam (Figure 2).

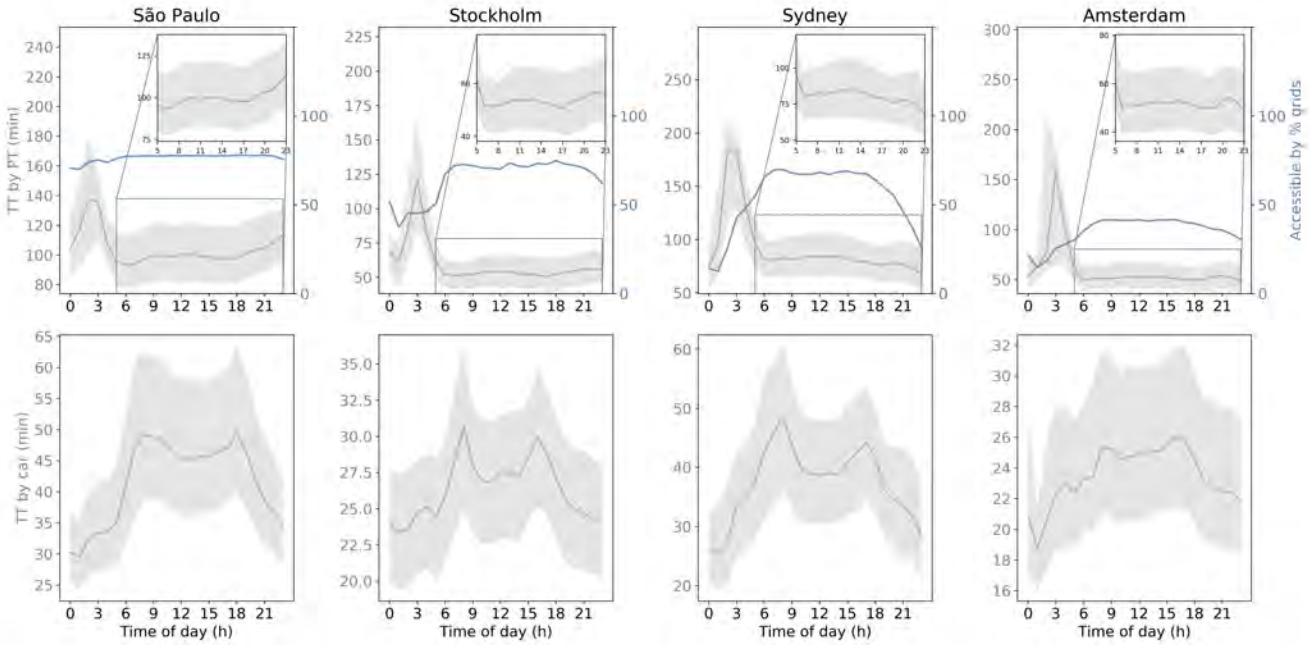

Figure 1. Temporal variations of travel time by public transit (upper row) and by car (bottom row) over the course of an average weekday. The travel time is the citywide average across departure locations, weighted by population density, of the average travel times from those locations. The shaded area indicates the range from the 25th to 75th percentile. Also shown is the percentage of grid cells accessible by PT by time of day. The inset figures are zoomed into the time period of from 05 hours to 23 hours to better show the variation of the travel time by PT.

The travel time ratio R in Figure 3 shows the spatial distribution of daily average travel time (see Methods) by PT compared with car use in contrast with the population density distribution. For each grid cell, taking PT takes around twice as long as using the car for all the study areas. In all four study areas, the spatial distribution of R reflects the spatial organisation of cities so that the advantage of car travel over PT tends to be smaller near city centres, in regions with higher population and activity densities, and closer to PT infrastructure.

Cross-regional comparison of the modal disparity of travel time

The temporal variation of R shown in Figure 4(a) suggests that the average travel time ratio is around 2 throughout most of the day, and the highest disparity between the two modes occurs between midnight and before dawn, when PT service is typically reduced or not running at all. The share of area that favours PT over car use is very small: 0.58%, 0.43%, 0.98% and 1.12% (daily average) or 0.61%, 0.46%, 1.14% and 1.14% (during peak hours) for São Paulo, Sydney, Stockholm, and Amsterdam, respectively. Figure 4(b) shows how the travel time ratio changes as travel time by car increases. It turns out, however, that PT outperforms cars on shorter trips. For trips with car travel times greater than 18 min, the disparity between the two modes slowly decreases.

Figure 5(a) shows the travel time by car and by PT as a function of travel distance (by car for the sake of comparison). Given the same travel distance, these cities have different travel times by car. São Paulo has the longest travel time by car for the same distance due to the high congestion level[24]. Within the range of 0 - 40 km, the three other cities show similar levels of travel time by car, with Amsterdam having the lowest. Despite similar sizes, São Paulo and Sydney display different patterns of travel time by car with increasing travel distance: Travel time in São Paulo increases with distance much more rapidly than in Sydney, presumably due to the high congestion levels. São Paulo also has the highest travel time with PT for the same distance, whereas the three other cities have similar travel times by PT for the same distance. In general, travel time by PT increases quickly for short-distance trips (0 - 20 km) but increases rather slowly for medium-distance trips (20 - 40 km). For long-distance trips ($>$ 40 km, in São Paulo and Sydney), the travel time once again becomes sensitive to the increased distance.

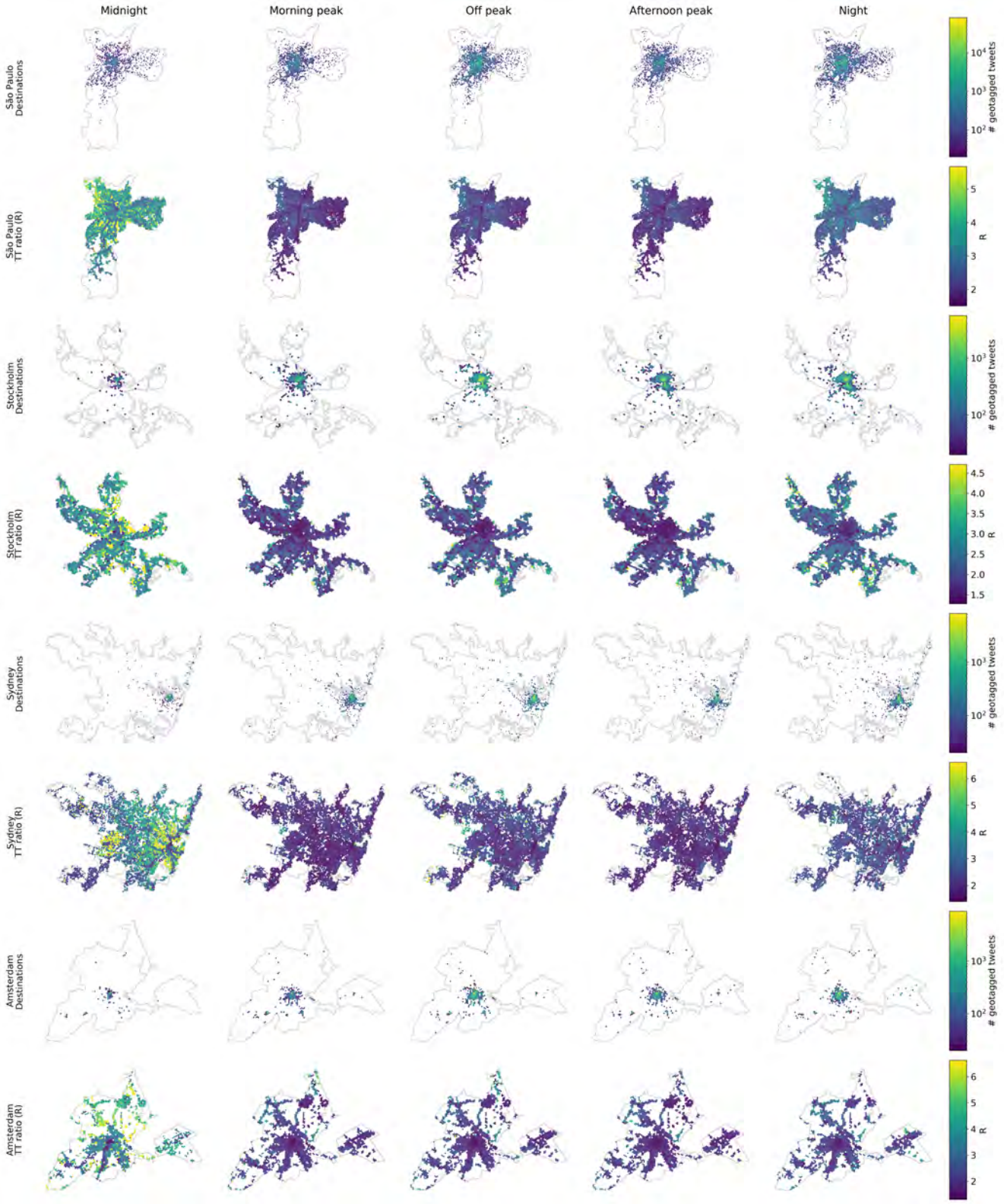

Figure 2. Spatiotemporal variation of travel time ratio (R) over the course of an average weekday. The travel time ratio R is the ratio of the average PT travel time to the average car travel time. For each city, the upper row shows the visited locations weighted by their number of geotagged tweets, and the bottom row shows the weighted average travel time to those locations within each time interval. Midnight = 0:00 - 7:00, Morning peak = 7:00 - 10:00, Off peak = 10:00 - 16:00, Afternoon peak = 16:00 - 19:00, Night = 19:00 - 0:00. Maps presented with different spatial scales for the sake of clarity.

Figure 5(b) shows that the population can reach the frequently visited locations in the study areas by car faster than by PT, with travel time rising quickly and with a long tail in São Paulo and Sydney (i.e. the 10% of the population who live farthest away from the destinations have an average travel time that is roughly 3 times that of the closest 10%). In Figure 5(c), R can be less than 1 (PT faster than car use) for distances < 3 km, but PT quickly loses the advantage as distances increase. Except for Stockholm, the cities show similar patterns when travel distances continue to increase: The disparity between PT and car travel time continues to increase until it reaches a maximum value at around 15 km, and then it starts to drop. This can be explained by the slower increase of travel time by PT than by car. In addition, as shown in Figure 5(d), population density and R are also correlated: The greater the population density, the lesser the disparity between PT and car travel times.

At the city-level, Table 1 summarises the performance of PT and car use in terms of travel time ratio and the aggregate travel speed. The mean value of the (unweighted) travel time ratio for the study areas is 2.3 (São Paulo), 2.1 (Stockholm), 2.2 (Sydney), and 2.2 (Amsterdam), where the smallest difference between the modes is observed in Stockholm. The travel time ratio R weighted by population density is 1.4, 2.5, 2.3, and 2.1 for São Paulo, Stockholm, Sydney, and Amsterdam, respectively, suggesting PT services in São Paulo and Amsterdam are more closely matched with where people live versus the PT services in Stockholm and Sydney, which are focused more on spatial coverage. Although Stockholm has the smallest disparity between PT and car, when the average travel time ratio is weighted by the population density, Stockholm becomes the city with the largest disparity. However, São Paulo shows the opposite trend after incorporating population density; it becomes the city with the smallest disparity between PT and car. For PT, the differences of (population weighted) speed are small. For São Paulo, the low driving speed suggests heavy traffic congestion, explaining why the disparity in time between PT and car is smallest there.

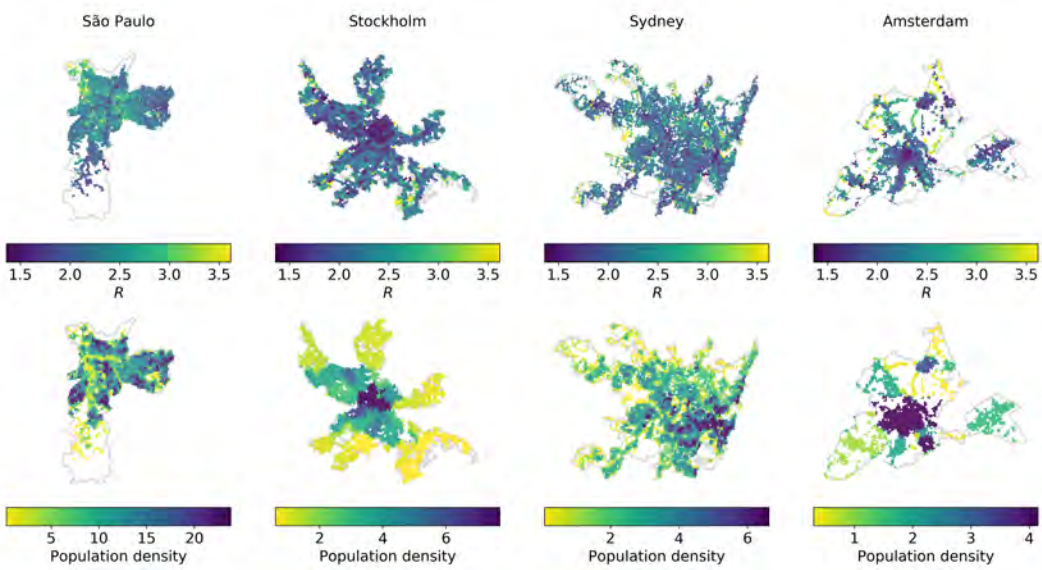

Figure 3. Spatial variation of travel time ratio (R) to frequently visited locations (top row) and in contrast with the population density distribution (bottom row). The value of R for each cell as the origin is the average value based on the 5% - 95% quantiles of travel times by PT and car in the time period between 05:00 and 23:00 weighted by the frequency of geotagged tweets in the destinations (same below). The warmer the colour, the greater the advantage of car use over PT. The unit of population density is 1000 persons per sq. km.

Discussion

This study used multiple large datasets to model travel time by car and by PT. HERE Traffic provides high-frequency data of the actual driving speed in the cities studied, with extensive coverage. We use GTFS data plus a trip planner as an "advanced" method[25] because GTFS data already takes into account congestion-related delays in route planning. The methodology in this study is repeatable and can scale up. Compared to previous research based on simple models using average speeds by mode, our methodology produces more realistic travel-time calculations, allowing for more reliable comparisons between the two modes, given that congestion can substantially alter driving times in urban settings when compared with theoretical estimates[25,37]. In addition to real-time road speed records and GTFS data, we introduce social media data to represent the locations of actual demand by time of day for an average weekday. The validity of using Twitter to represent the actual travel demand has been

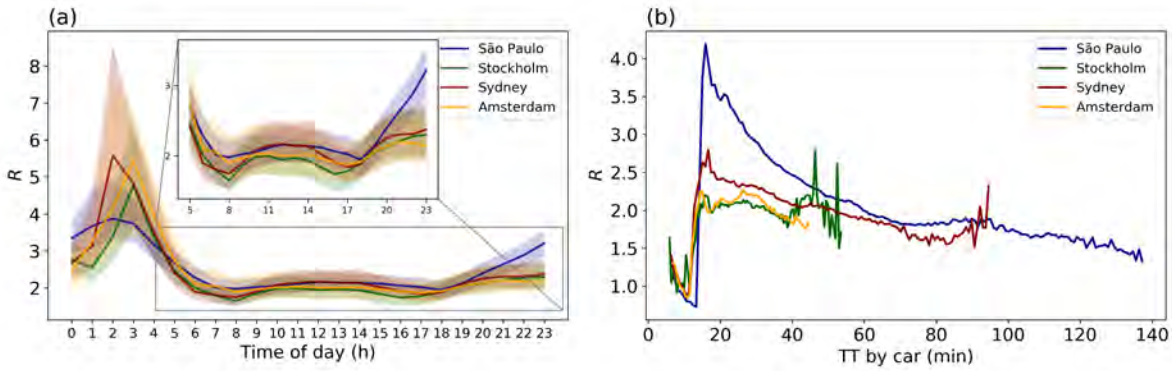

Figure 4. (a) Temporal variation of citywide average travel time ratio (R). The shaded area indicates the two mid quartiles. The insert zooms in on the period from 05 hours to 23 hours, to better show the temporal variation of R. (b) Travel time ratio (R) as a function of travel time by car.

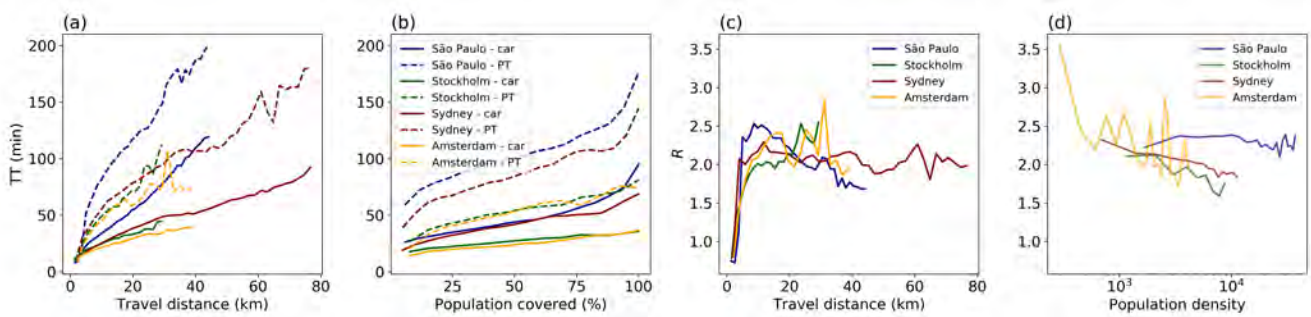

Figure 5. Travel time and travel time ratio vs. travel distance and population density. Travel time as a function of (a) travel distance and (b) the percentage of population reached. The relationship between travel time ratio (R) and (c) travel distance and (d) population density. The unit of population density is 1 person per sq.km.

thoroughly examined in previous studies[32,34].

The outcomes of the improved travel time calculations demonstrate the usefulness of applying large datasets in the framework developed in this study. The temporal and spatial variations in travel time for car and PT are described in detail for the four study areas (Figure 1 and in Figure 2). They show how the travel time for each mode changes by time of day for an average weekday and how travel time varies spatially in different cities. Future studies can zoom in and overlay infrastructure information to gain more detailed insights at the local level. This allows for urban planning policies to be better informed, especially in encouraging a mode shift from car to PT. While trips by PT take on average around twice as long as by car, this difference varies widely with location (Figure 2) and time of day (Figure 5). The travel time difference is consistent with previous studies, but this observation is now supported by better and more detailed data. In general, the area in the studied cities where PT can outperform car use is very small, despite there also being substantial areas surrounding PT lines where the disparity of travel time by car and PT is smaller than in the rest of the city.

We also present how the spatiotemporal variation of travel time differs between cities, and we further explore the impact of travel distance and population density on travel times across cities (Figure 5(a-b)). Travel times increase in different ways for cars and PT when the travel distance increases (Figure 5a). For both cars and PT, São Paulo shows significantly longer travel times compared with the other cities. To cover the same percentage of the total population, taking PT takes significantly longer time than driving a car (around twice, see Figure 5b). Moreover, for PT in the two big cities, São Paulo and Sydney, the final 10% of the population who live far away from the destinations have an average travel time that is roughly 3 times that of the first 10%.

In order to quantify the disparity between the two modes, the travel time ratio R is calculated and its relationship with travel distance and population density are demonstrated (Figure 5c-d), and we also calculate the city average of it (Table 2), see Methods. PT can outperform car use ($R <= 1$) for short distances (< 3 km), mainly during rush hour in Stockholm and Amsterdam, when the share of grid cells that favour PT can reach 0.8% in Amsterdam and 0.6% in Stockholm. This number is

City	Travel time ratio	Travel time ratio$_{pop}$[a]	Speed (km/h)		Speed$_{pop}$ (km/h)[b]	
			Car	PT	Car	PT
São Paulo	2.3	1.4	19.5	9.0	20.0	14.6
Stockholm	2.1	2.5	25.7	12.9	37.7	15.2
Sydney	2.2	2.3	33.8	16.5	31.4	14.1
Amsterdam	2.2	2.1	31.6	15.0	27.6	13.7

Table 1. Travel time ratio at the city level. [a-b] Average value weighted by population density in each grid cell. The travel time ratio at the city level is calculated based on the average value across all grid cells at all times of the day weighted by the frequency of geotagged tweets of the destinations (see Methods).

even smaller in Sydney and São Paulo. The results are consistent with the commonsense expectation that short travel distance, city centres, and the proximity to PT lines are situations in which PT can outperform car use. Similar patterns have been found in previous studies, e.g. of Greater Helsinki[25]. However, it is important to recognise that only a very small number of grid cells favour PT over car. At the city level (with grid cells weighted by population density), the lowest travel time ratio is observed in São Paulo, followed by Amsterdam, Sydney, and Stockholm in ascending order.

Travel time has been found to be the strongest predictor of mode choice[38], especially when PT is the alternative travel mode for car drivers[39]. A survey of car drivers conducted in Amsterdam showed that when the perceived travel time ratio dropped below 1.6, participants were willing to regularly make the trip by PT[39]. For the trips recorded in this survey, the average travel time by car was 35 min. However, our study shows that the travel time ratio is at best 2 for trips with car travel time of around 35 min (see Figure 4b). When asked "Could you also have made this trip by public transport?" a 35-minutes trip, prompts car drivers to say "yes, but rarely do". To get more car drivers to take PT, the travel time by PT needs to decrease further. However, the time elasticity of PT is greater than for cars, which means that to increase PT ridership, the marginal decrease of travel time is lower compared to for cars[40]. Travel time alone does not tell the whole story when encouraging a mode shift from car to PT, and PT being as fast as a car would not mean that everyone would prefer to take it[41]. Many other factors contribute to mode choice, e.g. comfort and price.

Conclusion

To assess the difference in travel time between car use and PT, we propose a computational framework incorporating new and large real-world data sets. We implement this framework in four cities and make a systematic comparison across them. The framework demonstrates its usefulness by revealing the travel time disparity between public transport and cars at a high spatial and temporal granularity enabling detailed and local-level explorations. On average, PT travel time is around twice as high as by car, confirming previous studies but with the most detailed real-world data. Importantly, we show how such disparity varies across space and time for an average weekday. The disparity tends to be much smaller near city centres and in the surroundings of PT lines. PT can outperform car use on average for short-distance travel (< 3 km) and during peak rush hour in Stockholm and Amsterdam. A systematic comparison between these two modes shows that the travel time disparity is surprisingly similar across cities: $R < 1$ for travel distances less than 3 km, then increases rapidly but quickly stabilises at around 2.

We will include more features from the geographical context to better compare private car with PT in future studies, including environmental performance, and especially the carbon footprint by mode.

Methods

The framework of data analysis is shown in Figure 6. We combine four big data sources in this study: HERE Traffic data, geotagged tweets, OSM road network, and GTFS data. Each city's study area is divided into hexagonal cells with a short diagonal of 500 m. All grid cells are represented by the latitude/longitude coordinates of their centroids. Geotagged tweets are used to derive the frequency of visited locations in each city. Some studies find good agreement, in general, between geotagged tweets and either traffic data[33] or travel-demand data[34]. In addition, geotagged tweets offer realistic activity (i.e. demand) locations[32] to avoid needing to estimate the travel time for all potential pairs of origin-destination in the city[25]. We classify locations determined to be visited frequently, i.e. those with more than 20 geotagged tweets over each hourly interval, as "destinations" and calculate the travel time by car and public transit from everywhere in the city to these destinations, see Travel Time Calculation.

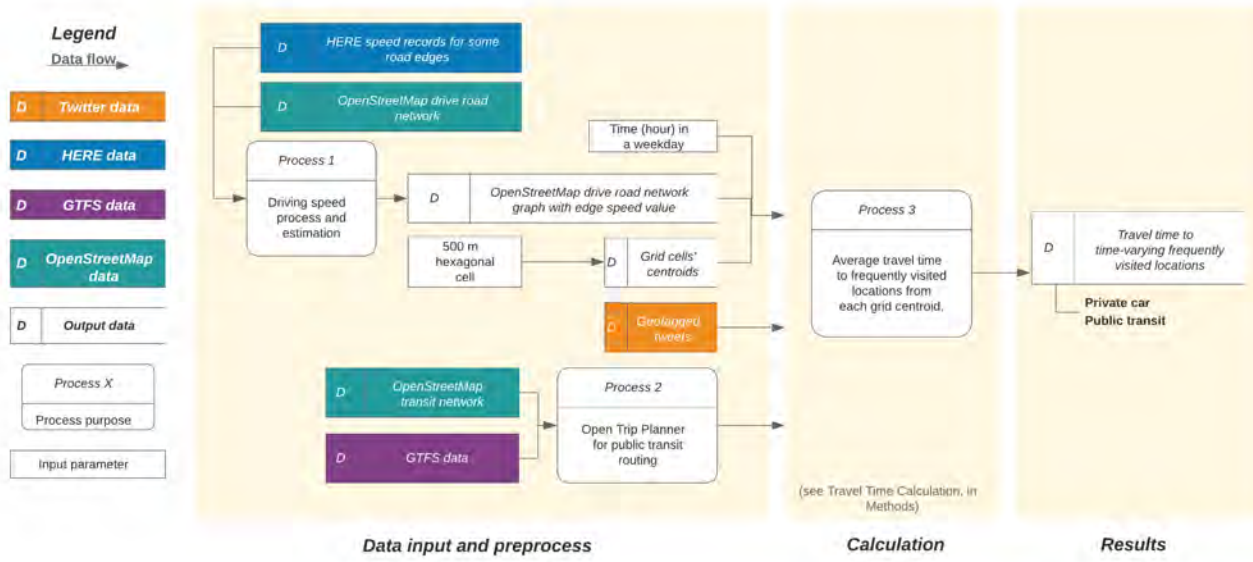

Figure 6. Flowchart of the analytic framework used for calculating the real-world travel time by car and PT.

Data and Preprocessing

HERE Traffic Data: Driving Speed

Data on driving speed were collected approximately every 5 minutes from January 1, 2018, to December 31, 2018. In this study, we focus on an average weekday. Therefore, we remove the data collected in January, February, July, August, and December to eliminate the impact of holiday seasons. Because of the variations in the timing of the sampling and varying latency in network delays, these samples are grouped into 15-minute time windows (96 windows per day), in which the measured traffic speeds in each bin are averaged. Roads are geographically represented by road segments as a sequence of edges. We calculate the hourly average speed for each road over weekdays. More details of the preprocessing method can be found in[24].

GTFS Data: Public Transit Timetables

A GTFS static dataset[iv] is a collection of text files consisting of all the information required to reproduce a transit agency's schedule, including the locations of stops and timing of all routes and vehicle trips. GTFS data were collected from various sources that are publicly available. São Paulo's GTFS data were obtained from OpenMobilityData[v] provided by SPTrans, Stockholm's from OpenMobilityData provided by Stockholm SL, Sydney's from the Open Data portal provided by the transportation department of New South Wales[vi], and Amsterdam's from the Open Data portal of 9292[vii].

OSM Data: Road Network for Driving

To calculate travel time by car, the road network is downloaded using osmnx[42] by specifying the road network type as "drive"and the area as per the study areas. The road network is then converted to a directed graph[42] and further converted into an igraph object[43]. The road network is also necessary for the selection of walking to and from PT stops. To calculate travel time by PT, the OSM data files are downloaded at the country level from GEOFABRIK [viii],and the cities' OSM data are further extracted using their corresponding poly files[ix].

Twitter Data: Time-varying Frequently Visited Locations

We use the Gnip database[x] purchased from Twitter to identify top non-commercial geotag users who geotagged their tweets most frequently during a six-month period (20 December 2015 - 20 June 2016) within the geographical bounding box of cities

[iv] https://gtfs.org/reference/static
[v] http://transitfeeds.com/
[vi] https://opendata.transport.nsw.gov.au/dataset/historical-gtfs-bundles-and-timetables
[vii] https://reisinformatiegroep.nl/ndovloket/datacollecties
[viii] http://download.geofabrik.de/
[ix] http://polygons.openstreetmap.fr/
[x] Gnip is a Twitter subsidiary which sells historical tweets in bulk, and provides access to the Firehose API. https://support.gnip.com/sources/twitter/

or countries of our study[44]. Top geotag users are selected by having at least 50 geotagged tweets per year. We extract these top geotag users' historical tweets occurring within the study areas using User Timeline API[xi]. The obtained geotagged tweets are further processed to represent the frequently visited areas in the cities. A more detailed description of the preprocessing methods of geotagged tweets and their validity in representing population-level demand can be found in our previous study[32].

For each hourly interval during weekdays, the frequently visited locations are represented by the centroids of each grid cell for which there are at least 20 geotagged tweets captured in. that hour. The total numbers of geotagged tweets and users for each city are summarised in Data Description, below.

Data Description

The summary statistics of the four cities are shown in Table 2. For São Paulo, the study area is the boundary of the municipality due to the lack of data for the metropolitan area. For the other cities, the study area is the functional urban area.

City	Scope	Area (km^2)	Population	GDP ($/capita)	Mode share (%) [a]			
					D	PT	W	B
São Paulo	Municipality	1,521	11,967,825	9,812	28	31	31	1
Stockholm	Urban area[b]	414	1,372,565	53,253	32	47	14	7
Sydney	UCL[c]	2,036	4,321,535	49,755	59	25	4	3
Amsterdam	Urban Area [d]	885	1,520,127	45,638	20	17	29	32

Table 2. Study cities' main characteristics and mode share (D = Driving, PT = Public transit, W = Walking, B = Biking). [a] Sources: São Paulo, Sydney, and Amsterdam in 2019 (Deloitte City Mobility, https://www2.deloitte.com/insights/mobility-index, Stockholm in 2011 (National Travel Survey). [b] Statistics Sweden - Open geodata for urban areas (https://www.scb.se). [c] Urban Centres and Localities defined by Australian Bureau of Statistics - Significant Urban Areas, Urban Centres and Localities (UCL), Section of State, July 2011 (https://www.abs.gov.au). [d] "Stadsgewesten" defined by Statistics Netherlands based on "Division of the Netherlands into 22 metropolitan agglomerations and urban areas" (2015, https://opendata.cbs.nl/statline/#/CBS/en/)

Basic information on the data sets collected for the four cities is shown in Table 3. The study areas and the spatial distribution of geotagged tweets are shown in Figure 7. In addition, we include population density data in the analysis[45]. Population density is used for visual comparison with the spatial distribution of travel time by different modes, and also used to weight the origin cells when aggregating the results across different grid cells for the travel time ratio (R) at the city level.

City	Grid cells #	HERE[a] (%)	GTFS (Date)	Twitter		
				User #	Geotag #	Time span[b]
São Paulo	7,423	55	2019-05-10	21,522	2,511,710	2010-09 - 2019-06
Stockholm	2,543	48	2019-05-15	5,344	310,609	2010-09 - 2019-03
Sydney	10,564	51	2018-05-14	11,707	608,699	2010-09 - 2019-05
Amsterdam	4,638	73	2018-05-28	9,424	284,281	2010-09 - 2019-04

Table 3. Basic information of the datasets in this study. [a] HERE indicator shows the share of real-time records. [b] Time span shows the time period from the earliest timestamp to the latest timestamp of geotagged tweets in the collected user timelines.

Travel Time Calculation

Origins, Destinations, and Travel Time

We calculate the travel time of fastest routes from the geographical centre of all grid cells to the frequently visited destinations indicated by geotagged tweets in the city. The destination cells vary depending on the time of day, aggregated on an hourly basis. During a certain time interval t (e.g., 8 am to 9 am), the average travel time of a given origin cell i, $T_{mode}(i,t)$, is defined as the mean value of the travel times from that origin to multiple destinations indexed as $j = 1,2,...,n(t)$ whose frequency of geotagged tweets $f(j,t)$ are used as weights:

$$T_{mode}(i,t) = \frac{\sum_{j=1}^{n(t)} [f(j,t) \cdot T_{mode}(i,j,t)]}{\sum_{j=1}^{n(t)} f(j,t)} \tag{1}$$

[xi] https://developer.twitter.com/en/docs/tweets/timelines/overview.html

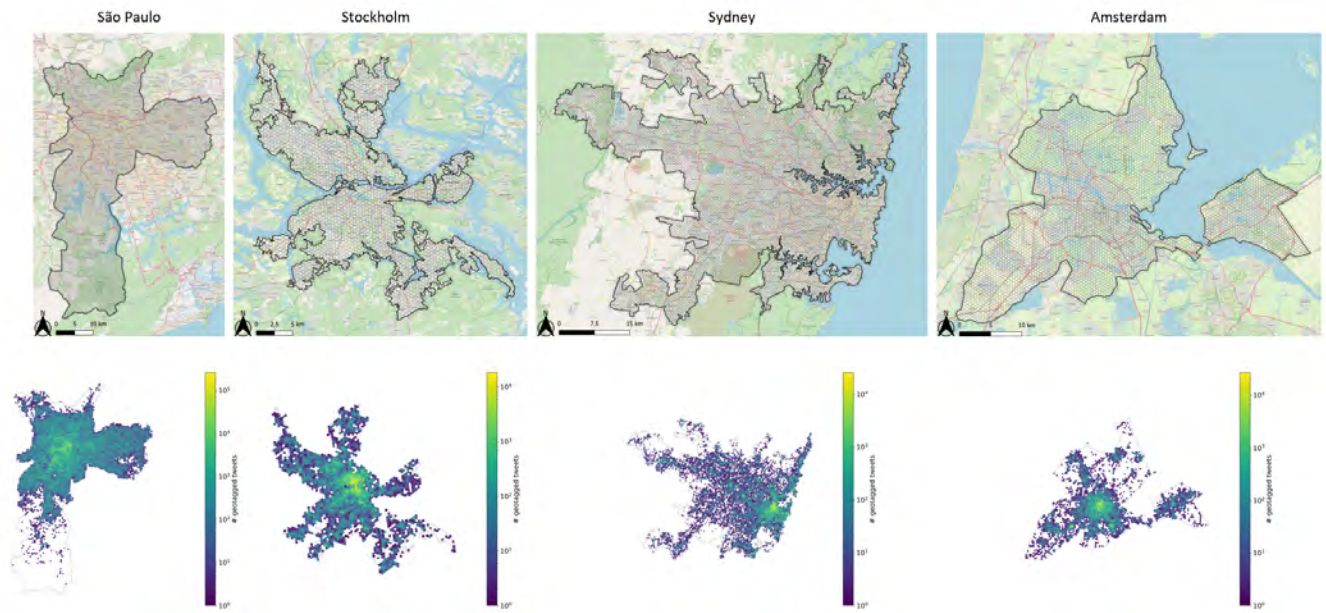

Figure 7. Studied cities: defined urban area (upper row) and spatial distribution of geotagged tweets (bottom row, all records are weekdays only).

where $T_{mode}(i,j,t)$ indicates the fastest travel time by a certain mode from origin cell i to destination cell j.

The average travel time over an average weekday is defined as:

$$T_{mode}(i) = \frac{\sum_{t=0}^{24}\sum_{j=1}^{n(t)}\left[f(j,t)\cdot T_{mode}(i,j,t)\right]}{\sum_{t=0}^{24}\sum_{j=1}^{n(t)} f(j,t)} \tag{2}$$

At the city level, the average travel time is calculated as:

$$T_{mode} = \frac{\sum_{t=0}^{23}\sum_{i=1}^{N(t)}\sum_{j=1}^{n(t)}\left[f(j,t)\cdot T_{mode}(i,j,t)\cdot Pop(i)\right]}{\sum_{t=0}^{23}\sum_{i=1}^{N(t)}\sum_{j=1}^{n(t)} f(j,t)\cdot Pop(i)} \tag{3}$$

where $N(t)$ represents the total number of hexagonal cells that can access the destinations at the given time frame and $Pop(i)$, when applied, indicates the population density of the origin cell i.

The detailed methods of calculating travel time by car $T_{car}(i,j,t)$ and by PT $T_{PT}(i,j,t)$ are shown in the next two sections.

Travel Time by Car Use

We calculate the travel time by car use through real-time road speed records from HERE Traffic and speed limits of downloaded drive road networks from OSM. Not every road in OSM is covered by HERE Traffic data. For roads that are not, the average driving speed is estimated based on the average speed for the same road type ("highway" tag in OSM) derived from HERE Traffic data, if applicable. For those roads that do not have real-time records or road-type average, the speed is estimated based on the speed limit in OSM if applicable. The speed of 30 km/h is assigned to any remaining road edges, which account for less than 0.05% of the total road edges.

Once the road speed is processed, a directed graph is constructed from the "drive" road network data, with the edges assigned a travel time based on the average speed and the length of that road edge. We calculate the travel time of each road edge with the departure time ranging from 00:00, at an hourly frequency, until 23:00 (included) on weekdays. The origin and destination are the nearest road nodes to the respective cells' centroids, and the impedance for routing is the travel time of road edge. The travel time from a given origin node to a given destination node is represented by the travel time of the fastest route identified using Dijkstra's algorithm[46] plus a random value between 5 and 10 minutes as the estimated time for parking following the method adopted in previous studies[47].

Travel Time by Public Transit

Travel time by PT is calculated based on GTFS data and OSM data on a normal weekday. Due to varying data availability across the study cities, we select the first Wednesday in May after the updating date of GTFS as shown in Table 3. Routing between a given origin-destination pair is conducted using an open-source multi-modal routing engine, OpenTripPlanner (OTP)[xii][48], similarly used in previous studies[21,49,50].

A trip by PT potentially consists of all available modes of public transportation (bus, tram, train, subway, etc.) and walking (speed = 1.4 m/s). For each pair of origin-destination, OTP finds the fastest door-to-door trip given a set departure time and the combination of transport modes available. The maximum walking distance is set to 800 m. Departure times are set to every half hour from 00:00 until 23:30 (included). A grid cell is marked as "not reachable by PT" if the fastest route between two locations does not meet the constraint of maximum walking distance of 800 m or if the trip takes longer than 240 minutes.

Relative Travel Time by Car Use vs. PT

We propose a travel time ratio calculated for grid cell i as the origin to the destination cell j quantifying the disparity between these two modes as shown below:

$$R(i,j,t) = T_{PT}(i,j,t) / T_{car}(i,j,t) \tag{4}$$

In this equation $R(i,j,t)$ is the ratio between travel time by public transit $T_{PT}(i,j,t)$ for an origin-destination pair (i) and the travel time by car $T_{PT}(i,j,t)$ for the same origin-destination pair ($i-j$). The higher the R, the less desirable PT is compared to the car. It is worth noting that we are comparing the empirical travel time by car with the scheduled PT travel time. While GTFS data normally takes recurrent congestion into account, scheduled transit services can be affected by unplanned factors such as traffic accidents and weather conditions, causing some deviations from the schedule[51].

When calculating the average values of R across different times or grid cells, the same aggregation method is applied as demonstrated in Equation 1-3.

References

1. Pucher, J. Public transportation. *The Geogr. Urban Transp.* **3**, 199–236 (2004).
2. Banister, D. Cities, mobility and climate change. *J. Transp. Geogr.* **19**, 1538–1546 (2011).
3. Rabl, A. & De Nazelle, A. Benefits of shift from car to active transport. *Transp. Policy* **19**, 121–131 (2012).
4. Edenhofer, O. *Climate change 2014: mitigation of climate change*, vol. 3 (Cambridge University Press, 2015).
5. Loy, T. So sollen bvg und s-bahn in zukunft fahren. *Der Tagesspiegel* (2019). URL https://www.tagesspiegel.de/berlin/nahverkehrsplan-so-sollen-bvg-und-s-bahn-in-zukunft-fahren/24038246.html.
6. Schafer, A. & Victor, D. G. The future mobility of the world population. *Transp. Res. Part A: Policy Pract.* **34**, 171–205 (2000).
7. Mallett, W. J. Trends in public transportation ridership: Implications for federal policy. Tech. Rep. (2018).
8. Alam, B., Nixon, H., Zhang, Q. *et al.* Investigating the determining factors for transit travel demand by bus mode in us metropolitan statistical areas. Tech. Rep., Mineta Transportation Institute (2015).
9. Cervero, R. Built environments and mode choice: toward a normative framework. *Transp. Res. Part D: Transp. Environ.* **7**, 265–284 (2002).
10. Ewing, R. & Cervero, R. Travel and the built environment: A meta-analysis. *J. Am. planning association* **76**, 265–294 (2010).
11. Beirão, G. & Cabral, J. S. Understanding attitudes towards public transport and private car: A qualitative study. *Transp. Policy* **14**, 478–489 (2007).
12. De Vos, J., Mokhtarian, P. L., Schwanen, T., Van Acker, V. & Witlox, F. Travel mode choice and travel satisfaction: bridging the gap between decision utility and experienced utility. *Transp.* **43**, 771–796 (2016).
13. Litman, T. Transit price elasticities and cross-elasticities. *J. Public Transp.* (2017).
14. Xin, Y., Fu, L. & Saccomanno, F. F. Assessing transit level of service along travel corridors: Case study using the transit capacity and quality of service manual. *Transp. Res. Rec.* **1927**, 258–267 (2005).

[xii] https://www.opentripplanner.org/

15. Redman, L., Friman, M., Gärling, T. & Hartig, T. Quality attributes of public transport that attract car users: A research review. *Transp. policy* **25**, 119–127 (2013).

16. Eriksson, L., Friman, M. & Gärling, T. Stated reasons for reducing work-commute by car. *Transp. Res. Part F: Traffic Psychol. Behav.* **11**, 427–433 (2008).

17. Renne, J. L. *Transit oriented development: making it happen* (Routledge, 2016).

18. Gil, J. & Read, S. Patterns of sustainable mobility and the structure of modality in the Randstad city-region. *A—Z ITU J. Fac. Archit.* **11**, 231–254 (2014).

19. Wang, P., Hunter, T., Bayen, A. M., Schechtner, K. & González, M. C. Understanding road usage patterns in urban areas. *Sci. reports* **2**, 1001 (2012).

20. Farber, S. & Fu, L. Dynamic public transit accessibility using travel time cubes: Comparing the effects of infrastructure (dis) investments over time. *Comput. Environ. Urban Syst.* **62**, 30–40 (2017).

21. Pereira, R. H. Future accessibility impacts of transport policy scenarios: equity and sensitivity to travel time thresholds for bus rapid transit expansion in rio de janeiro. *J. Transp. Geogr.* **74**, 321–332 (2019).

22. Alexander, L., Jiang, S., Murga, M. & González, M. C. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transp. research part c: emerging technologies* **58**, 240–250 (2015).

23. Xu, Y. & González, M. C. Collective benefits in traffic during mega events via the use of information technologies. *J. The Royal Soc. Interface* **14**, 20161041 (2017).

24. Verendel, V. & Yeh, S. Measuring traffic in cities through a large-scale online platform (in press). *J. Big Data Anal. Transp.* (2019).

25. Salonen, M. & Toivonen, T. Modelling travel time in urban networks: comparable measures for private car and public transport. *J. Transp. Geogr.* **31**, 143–153 (2013).

26. Durán-Hormazábal, E. & Tirachini, A. Estimation of travel time variability for cars, buses, metro and door-to-door public transport trips in santiago, chile. *Res. Transp. Econ.* **59**, 26–39 (2016).

27. Calegari, G. R., Celino, I. & Peroni, D. City data dating: Emerging affinities between diverse urban datasets. *Inf. Syst.* **57**, 223–240 (2016).

28. Luo, X. *et al.* Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: a case of shanghai, china. *J. Clean. Prod.* **142**, 926–935 (2017).

29. Pelletier, M.-P., Trépanier, M. & Morency, C. Smart card data use in public transit: A literature review. *Transp. Res. Part C: Emerg. Technol.* **19**, 557–568 (2011).

30. Lyons, G. Getting smart about urban mobility–aligning the paradigms of smart and sustainable. *Transp. Res. Part A: Policy Pract.* **115**, 4–14 (2018).

31. Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S. & Waller, T. S. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transp. Res. Part C: Emerg. Technol.* **75**, 197–211 (2017).

32. Liao, Y. & Yeh, S. Using geotagged tweets to assess human mobility: a comparison with travel survey and GPS log data (under review) (2019).

33. Ribeiro, A. I. J. T., Silva, T. H., Duarte-Figueiredo, F. & Loureiro, A. A. Studying traffic conditions by analyzing foursquare and instagram data. In *Proceedings of the 11th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, 17–24 (ACM, 2014).

34. Lee, J. H., Gao, S. & Goulias, K. G. Can twitter data be used to validate travel demand models. In *14th International Conference on Travel Behaviour Research* (2015).

35. Huang, W., Xu, S., Yan, Y. & Zipf, A. An exploration of the interaction between urban human activities and daily traffic conditions: A case study of toronto, canada. *Cities* **84**, 8–22 (2019).

36. Tenkanen, H., Saarsalmi, P., Järv, O., Salonen, M. & Toivonen, T. Health research needs more comprehensive accessibility measures: integrating time and transport modes from open data. *Int. J. Heal. Geogr.* **15**, 23 (2016).

37. Yiannakoulias, N., Bland, W. & Svenson, L. W. Estimating the effect of turn penalties and traffic congestion on measuring spatial accessibility to primary health care. *Appl. Geogr.* **39**, 172–182 (2013).

38. Frank, L., Bradley, M., Kavage, S., Chapman, J. & Lawton, T. K. Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transp.* **35**, 37–54 (2008).

39. van Exel, N. J. A. & Rietveld, P. Perceptions of public transport travel time and their effect on choice-sets among car drivers. *J. Transp. Land Use* **2** (2010).

40. Varotto, S. F., Glerum, A., Stathopoulos, A., Bierlaire, M. & Longo, G. Mitigating the impact of errors in travel time reporting on mode choice modelling. *J. transport geography* **62**, 236–246 (2017).

41. Schäfer, A. W. Long-term trends in domestic us passenger travel: the past 110 years and the next 90. *Transp.* **44**, 293–310 (2017).

42. Boeing, G. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Syst.* **65**, 126–139 (2017).

43. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **Complex Systems**, 1695 (2006). URL http://igraph.org.

44. Jeuken, G. S. *Using Big Data for Human Mobility Patterns – Examining how Twitter data can be used in the study of human movement across space*. Master's thesis (2017).

45. Center for International Earth Science Information Network CIESIN Columbia University, . Gridded population of the world, version 4 (gpwv4): Population count, revision 11 (2018). URL https://doi.org/10.7927/H4JW8BX5.

46. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271 (1959).

47. Forum, I. T. Benchmarking accessibility in cities (2019). URL https://www.oecd-ilibrary.org/content/paper/4b1f722b-en. DOI https://doi.org/https://doi.org/10.1787/4b1f722b-en.

48. OpenTripPlanner developers group. Opentripplanner. https://github.com/opentripplanner/OpenTripPlanner (2019).

49. Liebig, T., Piatkowski, N., Bockermann, C. & Morik, K. Dynamic route planning with real-time traffic predictions. *Inf. Syst.* **64**, 258–265 (2017).

50. Stewart, A. F. Mapping transit accessibility: Possibilities for public participation. *Transp. Res. Part A: Policy Pract.* **104**, 150–166 (2017).

51. Mazloumi, E., Currie, G. & Rose, G. Using GPS data to gain insight into public transport travel time variability. *J. Transp. Eng.* **136**, 623–631 (2009).

Acknowledgements

This research is funded by the Swedish Research Council Formas (Project Number 2016-1326 for YL), and by the Chalmers Area of Advance Building Futures (for JG).

Author Contributions

Yuan Liao, Jorge Gil, and Sonia Yeh conceptualised the study. Yuan Liao and Vilhelm Verendel preprocessed the data. Yuan Liao, Jorge Gil, and Rafael H. M. Pereira contributed to data processing and analysis. All authors wrote the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/srep. **Competing financial interests:** The authors declare no competing financial interests.