Analysing Urban Transport Using Synthetic Journeys

Marcin Luckner[1][0000−0001−7015−2956], Przemysław Wrona[1][0000−0002−5479−4489], Maciej Grzenda[1][0000−0002−5440−4954], Agnieszka Łysak[2][0000−0002−1439−6355]

[1] Warsaw University of Technology, Faculty of Mathematics and Information Science, ul. Koszykowa 75, 00-662 Warsaw, Poland
marcin.luckner,przemyslaw.wrona,maciej.grzenda@pw.edu.pl

[2] Jan Kochanowski University, Institute of Physics, ul. Uniwersytecka 7, 25-406 Kielce, Poland agnieszka.lysak@ujk.edu.pl

Abstract. Travel mode choice models make it possible to learn under what conditions people decide to use different means of transport. Typically, such models are based on real trip records provided by respondents, e.g. city inhabitants. However, the question arises of how to scale the insights from an inevitably limited number of trips described in their travel diaries to entire cities.

To address the limited availability of real trip records, we propose the Urban Journey System integrating big data platforms, analytic engines, and synthetic data generators for urban transport analysis. First of all, the system makes it possible to generate random synthetic journeys linking origin and destination pairs by producing location pairs using an input probability distribution. For each synthetic journey, the system calculates candidate routes for different travel modes (car, public transport (PT), cycling, and walking). Next, the system calculates Level of Service (LOS) attributes such as travel duration, waiting time and distances involved, assuming both planned and real behaviour of the transport system. This allows us to compare travel parameters for planned and real transits.

We validate the system with spatial, schedule and GPS data from the City of Warsaw. We analyse LOS attributes and underlying vehicle trajectories over time to estimate spatio-temporal distributions of features such as travel duration, and number of transfers. We extend this analysis by referring to the travel mode choice model developed for the city.

Keywords: Travel mode choice · synthetic journeys · public transport

1 Introduction

Modern cities have typically been designed and built with the primary focus on the needs of car drivers [10]. Planning concepts, such as the 15-minute city, aim to minimise car usage by ensuring access to critical urban facilities within walking distance. Another approach is to promote less energy consumption and pollution-emitting means of transport [2]. However, the proposed solutions can be difficult to implement in the existing urban infrastructure [8]. Therefore, it is necessary

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

to develop tools that enable data-driven decisions for urban development. These include vehicle trajectory and LOS analysis tools that can give insights into traffic and street congestion which are useful for routing and transportation planning [14].

Travel mode choice (TMC) modelling [4,5,6] is vital to understanding under what conditions people decide to use different means of transport. In particular, it helps us to understand what makes people use (or not use) public transport. However, while TMC models predict whether for a trip of interest a car or another means of transport is likely to be used, identifying the spatial distribution of mode choices in urban areas is difficult. This is because the number of trips collected in surveys is inevitably limited, and increasing this number is expensive. Moreover, as not everyone is equally likely to share their data, increasing the size of the representative sample of real journeys in the areas of interest is additionally difficult.

Motivated by these needs, we propose the Urban Journey System (UJS) integrating Apache big data platforms with spatial data analytic engines based on OpenTripPlanner (OTP). The system combines open-source platforms with the newly proposed JourneyGenerator, JourneyDescriber, and JourneyAnalyser modules. JourneyGenerator generates synthetic journeys which are used to create public and individual transport trajectories. In this way, an arbitrarily large number of journeys can be obtained. To obtain representative journey origin and destination pairs, locations are generated using an input probability distribution, such as the probability distribution of journey endpoints based on time-dependent transport model demand matrices.

Next, JourneyDescriber calculates candidate routes for each synthetic journey. These include routes for car, walking, cycling, and PT using OTP instances provided with planned and real timetables in the form of General Transit Feed Specification (GTFS) files. While planned GTFS files are obtained from transport authorities, real GTFS files are developed by our system using a real-time location stream of public transport vehicles processed inter alia by a module based on Apache Flink. JourneyDescriber also calculates LOS attributes for various travel modes. Finally, the JourneyAnalyser produces an interactive HTML report on the generated journeys and compares LOS attributes.

We validate the system with the results obtained for the City of Warsaw, Poland. The GTFS feed and real-time PT location stream were processed to calculate scheduled and real PT networks. Hence, we use a selection of 399 planned and real daily public transport schedules already collected by the system. Journeys are generated based on the origin-destination hourly demand matrices of the transport model. In our analysis, we pay particular attention to journey attributes having a key impact on travel mode choices.

The remainder of this work is organised as follows. In Sect. 2 we analyse related works. This is followed by a summary of the system in Sect. 3, including an overview of its implementation. Next, results obtained for the City of Warsaw are analysed in Sect. 4. Finally, conclusions are made in Sect. 5.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

2 Related works

Several other works have focused on the spatial distribution of the choice of means of transport and its LOS. The applied approach to estimating their spatial distribution depended on the available data. Chia et al. explored the relationship between the spatial distribution of transfer location and the transit service's attractiveness in Brisbane [3]. The study was limited to services operated with a passenger card because the analysis was partially based on data the card operator system collected. The rest of the data came from a travel survey.

Yousefzadeh Barri et al. explored data from an extensive household travel survey in the Toronto region, including a one-day household travel diary [15]. They used statistical and machine learning (ML) models to predict newly generated transit trips by a low-income carless group after improving job accessibility. The study was limited to the use of public transport. Rocha et al. combined genetic algorithms and geostatistical methods to forecast travel demand variables and, as a result, the distribution of car trip rates in the São Paulo Metropolitan Area [11]. The data used in the study came from the Origin-Destination (OD) survey. Regarding means of transport, the study was limited to cars.

Tenkanen et al. analyzed different travel modes in the Helsinki region [13]. The analysis included door-to-door walking, cycling, driving and transit journeys. Centroids of statistical grid cells were used as origin and destination points for the calculation of the journeys. Their work is focused on distance and duration. Other LOS attributes were not analyzed. However, some of them were estimated to obtain the total travel time, e.g. PT walking times to and from the nearest stops were estimated based on Euclidean distances.

In [1], public transport time inaccuracy and variability were addressed. The duration of door-to-door PT journey and its components e.g. waiting and in-vehicle time were estimated using both scheduled timetables and actual timetables determined based on past GPS traces of PT vehicles. Both the origin and destination of each generated journey were based on centroids of the hexagonal spatial index developed by Uber with the area of each hexagon of 0.1km^2. Geolocated jobs were used to enable spatial analysis of employment accessibility. LOS attributes of other than PT transport modes were not considered. Demand matrices were not used to vary demand for transportation between zones. Travel time reliability was recently addressed also in [16], where a proposal for formulas quantifying PT competitiveness compared to cars based on maximising entropy value to obtain more dispersed competitiveness was made. Maps of areas of Hangzhou, China with low and high competitiveness according to these formulas were obtained.

As observed in [6], travel mode choice models most frequently rely on survey data only. This is even though the LOS attributes documenting the choices faced by a traveller are also considered to be important [5]. Among the attempts to calculate features quantifying trip characteristics under different travel modes, considering exact point coordinates, the study developed for London can be mentioned [6]. The LOS attributes in the work included durations of walking, cycling, interchanges and the whole PT route, and were calculated for real trips

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

only. No attempts to generate trip endpoints of representative trips were made. Furthermore, the spatial variability of the values of LOS features was not considered [6]. The inevitably limited number of real trips, compared to the London city area, illustrates the challenges caused by the use of real trip data.

In this work, we aim to go beyond these studies by generating and analysing an arbitrary number of detailed trips in urban areas, linking points generated based on probability distributions calculated with data from a transport model. Our system enables spatial analysis of multiple trip features under four different travel modes rather than the use of PT and car only. We consider inter alia features found relevant for TMC models built with ML methods.

3 System Overview

The objective of the UJS system is to enable large-scale analysis of journeys in urban areas. The high-level architecture of the system including its core components is presented in Fig. 1. Let us note that both JourneyGenerator and JourneyDescriber were designed to be a part of the Use4IoT architecture [7].

The PT schedule data in GTFS format for each day is developed based on data downloaded from transport operators. This provides planned daily timetables. However, disruptions such as delays and cancellations sometimes occur in public transport. Hence, real GTFS is created using a real-time location stream of public transport vehicles obtained from the API of public transport entities and processed inter alia by Apache Flink. In this way, real daily timetables in GTFS format are obtained. This makes it possible to construct a real transit network. Hence, for example, delays causing possibly missed transfers can be considered when calculating LOS attributes for a day of interest, based on the real behaviour of the PT system. This enables analysis of the actual experience of using PT.

3.1 JourneyGenerator

Planning an urban transport system is a complex task that relies on travel demand patterns. To model such demand, the urban area is frequently divided into disjoint transport zones. A travel demand (TD) matrix is a square matrix which shows the expected number of passengers moving between each combination of zones. The columns represent the origin zones, while the rows represent the destination zones of the city. In our approach, we use a potentially different matrix for each hour of the day. Hence, we consider up to 24 demand matrices for a demand scenario. One scenario can denote, e.g. real demand observed currently or hypothetical demand expected in 5 years during working days. Each matrix can come from the transport model of the city or be generated to reflect, e.g. the number and location of children travelling to schools.

TD matrices are used in JourneyGenerator to generate synthetic journeys, as shown in Alg. 1. We use a demand matrix to generate a random zone pair in line 4 of the algorithm. Although any zone combination is possible, the likelihood of

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

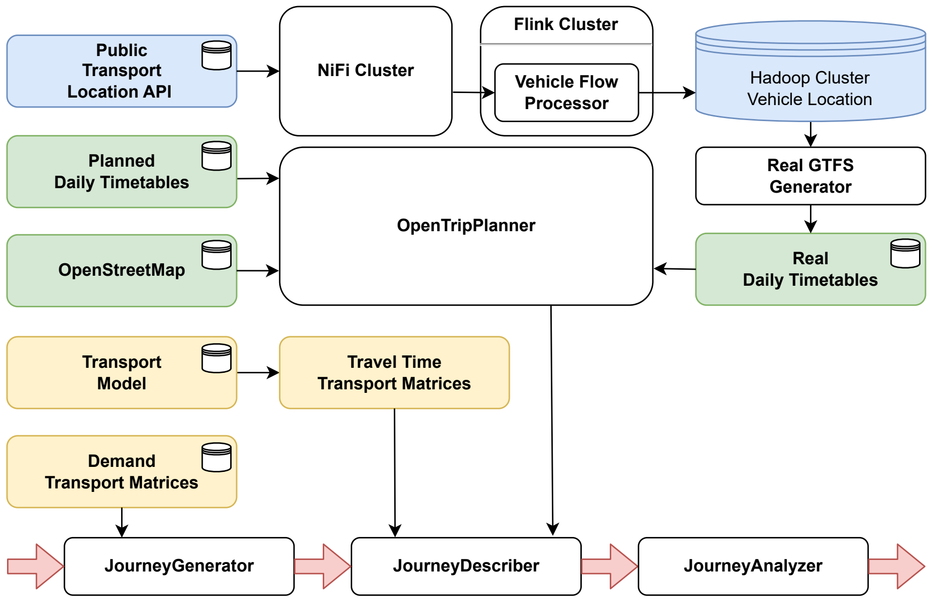


Fig. 1: High-level architecture of UJS system

selecting each zone varies. A zone pair and an hour with higher passenger traffic between the zones are more likely to be selected. After randomly selecting an origin zone and a destination zone, in lines 5 and 6 we select random addresses from the lists of address points in each zone area. The probability of selecting an address is proportional to its weight. This makes it possible, e.g. to make journeys from buildings populated by a large number of residents more likely to be generated. Finally, an exact time during the one hour is randomly generated.

3.2 JourneyDescriber

For each synthetic journey produced by JourneyGenerator, JourneyDescriber calculates trajectories likely to be used with different travel modes and LOS attributes for each travel mode. Estimating a travel path and calculation of its LOS attributes such as duration and distance is carried out using OTP[3], a multimodal trip planning module relying on OpenStreetMap[4] (OSM) [9]. OSM provides information on transport infrastructure such as the street network and the location of public transport stops.

OTP calculates private car routes and estimates trip duration using OSM data about the street network, taking into consideration traffic regulations and obstacles such as traffic lights, road types, and crossroads limiting the estimated

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

[4] Map data copyrighted by OpenStreetMap contributors and available from https://www.openstreetmap.org

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

Algorithm 1 The generation of journeys

Input:
int N - the number of journeys to generate
int[][][] *matrices* - a table of H two-dimensional matrices representing the number of passengers moving between zones $(i, j), i, j \in \{1, \ldots, Z\}$, one matrix per one hour time slot $h \in \{h_1, \ldots, h_H\}$.
$A[]$ *addresses* - list of address points to consider in each zone $i \in \{1, \ldots, Z\}$

procedure GenerateRandomJourneys(N, matrix, addresses)
  $journeys \leftarrow emptyList()$
  for $k \leftarrow 1$ to N **do**
    $(C_{k_{start}}, C_{k_{end}}, h) \leftarrow GetRandomZonePair(matrices)$
    $A_{k_{start}} \leftarrow GetRandomAddress(C_{k_{start}}, addresses)$
    $A_{k_{end}} \leftarrow GetRandomAddress(C_{k_{end}}, addresses)$
    $journeys.add(A_{k_{start}}, A_{k_{end}}, randomTime(h))$
  end for
  $return(journeys)$
end procedure

car speed. Based on travel time matrices from a transport model, JourneyDescriber also calculates the travel time by car under the expected street congestion for a given zone pair and time of the day.

In the case of public transport, the routes are calculated between public transport stops, and the waiting and walking times are added to the total trip duration. The calculated distance includes walking from the trip origin to the first stop, from the last stop to the destination, and the potential distance covered during transfers in multimodal travel.

For a private car, a single route is calculated. For PT, a set of routes is created that consists of all connections that start up to 5 minutes before and 10 minutes after the given journey starting time. Sample car and PT routes calculated for the same input synthetic journey are presented in Fig. 2.

For every synthetic journey, JourneyDescriber makes requests to the OTP instance(s). Each instance is configured with planned or real daily timetables. OTP provides JourneyDescriber with data about possible routes for car, PT, walking and cycling, and LOS attributes such as travel duration for each of these modes. Importantly, LOS features can be calculated using both planned timetables and real timetables developed by the UJS system based on GPS traces of public transport vehicles. Thus, LOS features documenting planned connections can be compared with LOS features quantifying connections which were feasible in the past. In particular, the impact of delays on missed connections is reflected in the LOS features developed based on real timetables.

In the OTP responses, for requests to OTP instances based on both planned and real timetables, we also receive information inter alia on potential PT connections within a 15-minute timeframe, the duration of each connection, and the number of transfers required by each of them. Data on possibly many PT connections per journey is aggregated by JourneyDescriber to determine LOS

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

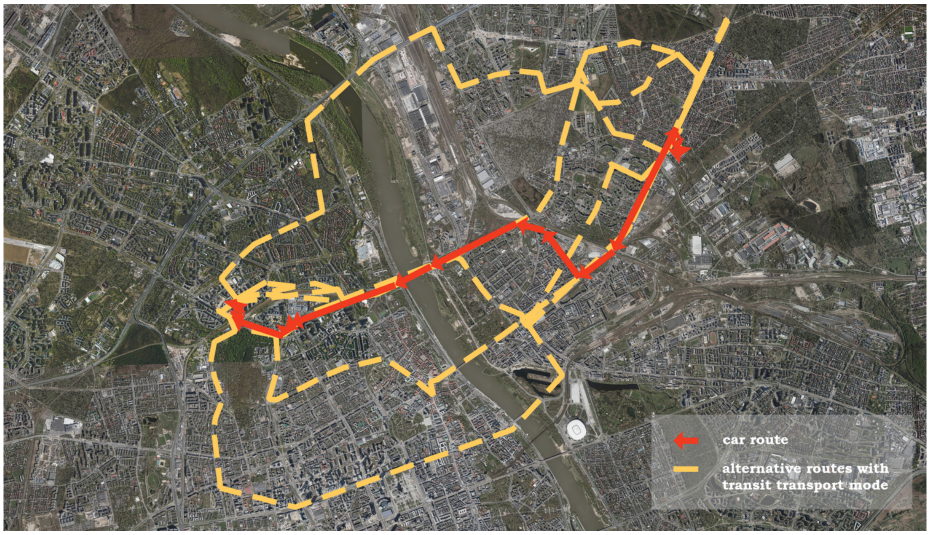


Fig. 2: Example car route and its PT alternatives

attribute values such as minimum and average travel time and the number of possible connections within a 15-minute window from the start of the journey. Finally, every journey with its LOS attributes and trajectories is included in the output list of synthetic journeys.

3.3 JourneyAnalyser

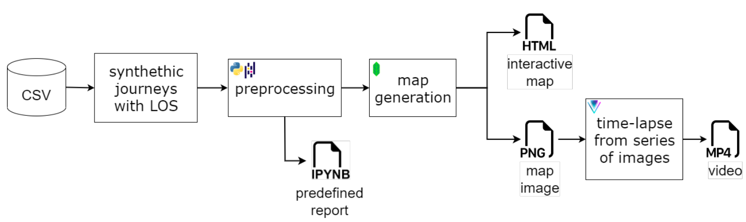
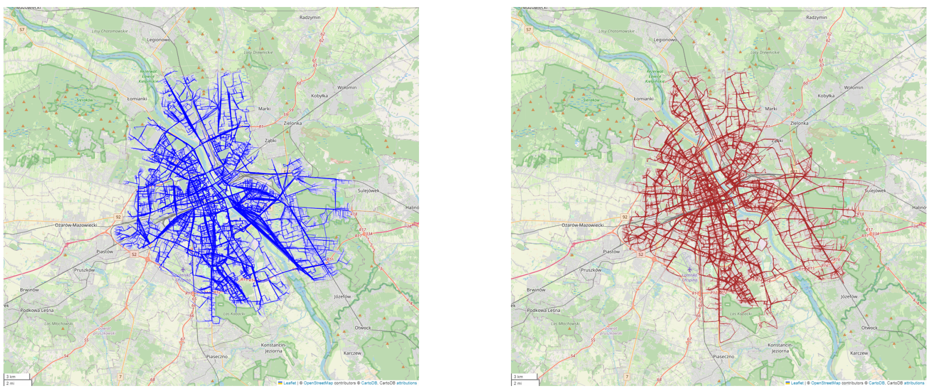


Fig. 3: The workflow of JourneyAnalyser

Fig. 3 shows the workflow of JourneyAnalyser – a module implemented in Python that generates the final analysis. Input journey records in CSV format with embedded trajectories are filtered to develop maps and plots, e.g. for specific hours and means of transport. Data is also transformed to extract individual car and PT trajectories for each journey. The Folium library generates interactive maps from preprocessed data and OSM. The maps are exported in PNG format

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

for coarse analysis and in HTML format for more geospatial insight. Fig. 4 presents sample maps for car (Fig. 4a) and PT (Fig. 4b) connections. A time-lapse video from the PNG maps can be generated for the period of interest to better understand the geospatial-temporal data. The preprocessed data is also the source for predefined reports implemented as a Jupyter Notebook. The reports compare global LOS for means of transport, timetables or hours, e.g. by generating empirical cumulative distribution function (ECDF) plots for LOS attributes such as the ones shown in Fig. 5 and Fig. 6.



(a) Routes of cars. 6 am to 12 pm

(b) Connections by PT. 6 am to 12 pm

Fig. 4: Sample interactive maps produced by JourneyAnalyser. SYNTH_WAW_EQW data. Background: [9].

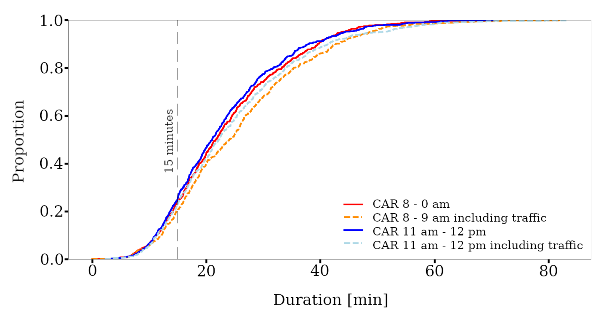


Fig. 5: ECDFs of car trip duration not considering street congestion and considering congestion. Selected hours. SYNTH_WAW_EQW data.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

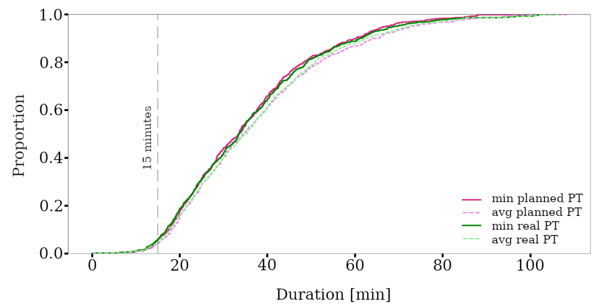


Fig. 6: ECDFs of PT trip duration, including minimum and average travel time under planned and real timetables. All considered hours. `SYNTH_WAW_EQW` data.

3.4 Implementation

The system was implemented at a central unit with 2xLenovo P/N BF78 2.65GHz and 48 cores, 1536 GB RAM and 64TB mass storage, which provided the basis for Apache NiFi, Apache Hadoop, Apache Flink, six OTP instances, and R and Python environments. It serves UJS needs by inter alia collecting planned timetables and GPS traces of PT vehicles, and calculating real timetables.

4 Results

4.1 Aggregation of journey features into frequent routes and distribution functions

The system was validated using real data from the City of Warsaw. First, public transport APIs were used to download raw data and prepare planned and real daily timetables for buses and trams. For the metro, GTFS files were calculated using metro frequency information. The Warsaw transport model was used to provide travel time transport matrices for cars and OTP was fed with the Warsaw infrastructure spatial data from OSM.

Two synthetic data sets were developed with UJS. In both cases, real demand matrices for working days from the transport model were used as input for JourneyGenerator. In the first case, the addresses considered included all available addresses in the City of Warsaw, sampled with equal weights. 4,000 journeys in the period of 6 AM to 12 PM based on these weights were generated with Alg. 1. LOS attributes were calculated using both planned schedules and real schedules. This was done by submitting requests to two different OTP instances in which planned and real Warsaw transport networks were configured. In this way, the `SYNTH_WAW_EQW` data set was developed.

Data sets such as the `SYNTH_WAW_EQW` data set can be used for both calculating distributions of individual features and visualising routes of synthetic trips on maps. Examples of interactive maps generated by JourneyAnalyser are given in Fig. 4. Fig. 4a identifies segments of traffic infrastructure heavily used

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

by individual means of transport in morning hours (between 6 am and 12 pm). Traffic congestion is visible on the east bank of the Vistula River in the south part of the city. Studies on air pollution [12] can be complemented by this kind of analysis.

Similarly, Fig. 4b shows the quickest PT connections which could be used for journeys present in the `SYNTH_WAW_EQW` data set. Public transport in Warsaw mainly consists of low-emission vehicles (buses and electric trams). Therefore, this visualisation can be used to identify areas without good direct connections with the city centre. Such areas are observed in the north-west outskirts of the city along the Vistula river, splitting the city into two parts. In this area, there are more direct car routes than PT connections.

These results can also be analysed statistically, e.g. aggregated using an empirical cumulative distribution function (ECDF). Fig. 5 and Fig. 6 show the distribution of travel duration. Fig. 5 illustrates the differences in the distribution of car travel duration during and after morning rush hours. The plots compare the duration calculated with and without considering traffic congestion. Let us note that not only congestion but also travel origin-destination patterns are different depending on the time of the day.

Similarly, Fig. 6 compares PT travel duration calculated using planned and real daily timetables. While the planned travel duration is calculated using planned timetables, the real travel duration is calculated using data from GPS sensors in the vehicles. Because the duration for PT is calculated as a statistic from available connections for each trip, minimal and average travel times are compared. The plots show that the duration difference between planned and real connections is higher for longer journeys, which frequently include transfers between lines and are more susceptible to delays of individual vehicles. Both Fig. 5 and 6 illustrate how data produced by UJS can be used for in-depth analysis of LOS attributes under different conditions.

4.2 Analysing spatial distribution of the level of service features

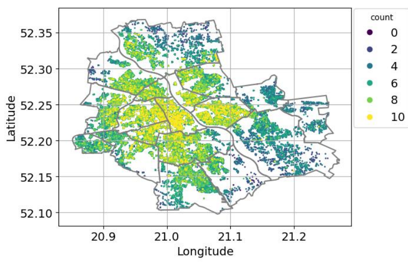
The `SYNTH_WAW_EQW` data set shows that aggregates of trips and ECDFs can be developed with moderately sized data sets. However, a few thousand trips are not sufficient to generate high-density plots showing spatial distribution of LOS.

Hence, the second data set used as an illustration in this study was populated with 100,000 journey records. This time the addresses considered included all available addresses in the City of Warsaw, sampled with weights proportional to the number of inhabitants of an address. As before, the journeys were randomly generated with Alg. 1. LOS attributes were calculated using planned schedules. In this way, the `SYNTH_WAW_INHW` data set was developed. Trips were generated for random times for hours between 6 AM and 8 PM. As discussed in Sect. 3.1, the probability of selecting an hour depended on the overall number of journeys for this hour according to the demand matrices. In this way, journeys during rush hours were more likely to be produced.

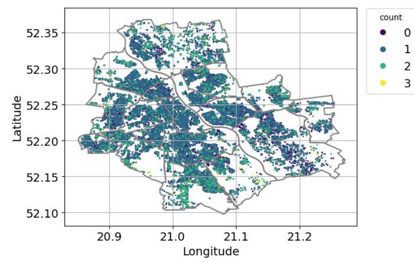
Maps showing the distribution of selected LOS features are provided in Fig. 7. Each point represents the location of the origin of the synthetic trips present in

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

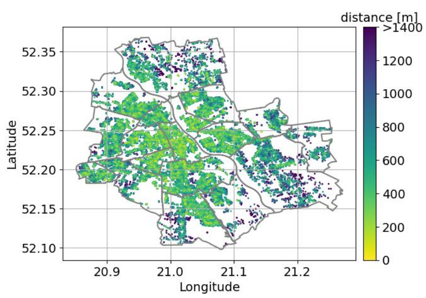
the data. It can be observed that the density of points varies greatly, which corresponds to the fact that trips from some locations, such as forest areas, are far less likely during the working days for which the data set was generated.



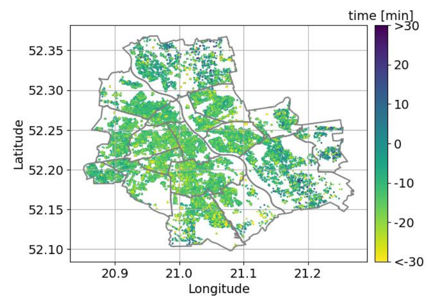
(a) Number of available PT connections



(b) Minimum number of required transfers



(c) Minimum walking distances required to use PT connections



(d) Difference of travel duration between bicycle and PT

Fig. 7: Distributions of sample LOS values. `SYNTH_WAW_INHW` data.

Fig. 7a shows the number of available PT connections to a destination within 15min i.e. in the period $[t(j) - 5min, t(j) + 10min]$, where $t(j)$ denotes the randomly assigned start time of journey j. In the city centre, 10 or more connections are frequently available during such periods, as travellers may rely on multiple trams and buses travelling through major transportation hubs. However, it can be observed that for some areas from which many trips are likely to be initiated such as the eastern part of the city, the number of available connections is significantly lower. Similarly, in such areas, as shown in Fig. 7b, even the best connection may require one or more transfers.

Fig. 7c shows that the overall walking distance needed by travellers in some areas of the city to use a suitable PT connection varies greatly. Importantly, when generating LOS features, UJS considers feasible routes, i.e. in this case walking paths resulting from street and pavement networks rather than Euclidean distances. In less densely populated areas the density of feasible walking paths is likely to be much lower, resulting in a major difference between the Euclidean

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

distance and the actual length of the walking route. Finally, Fig. 7d shows the values calculated by subtracting from travel duration by bicycle the duration of the quickest connection with public transport.

4.3 Adding spatial aspects to travel mode choice modelling

Let us note one more use for the data generated by the system. Municipalities struggle to reduce air pollution and foster sustainable mobility. Hence, the question arises of how to analyse the potential for environmentally-friendly travel mode choices. In Fig. 8, we provide a TMC model developed for the City of Warsaw, taking the form of a decision tree and pruned to include key top-level splits only [5]. Importantly, the model was developed using both real survey data including trip diaries and an extensive set of LOS features, as suggested in [6].

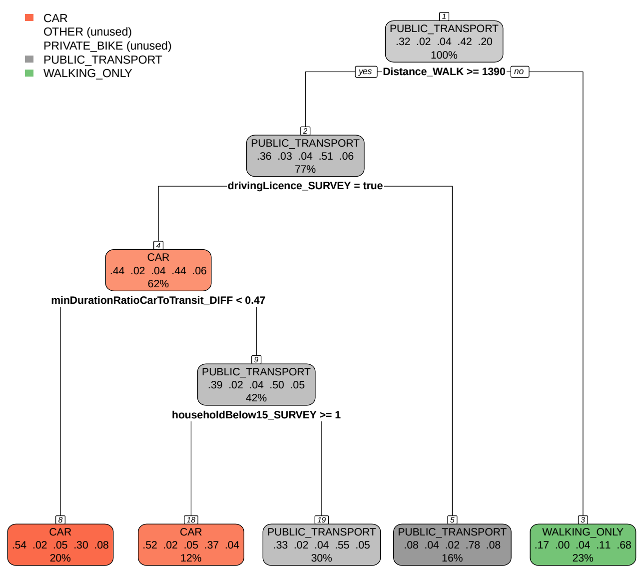


Fig. 8: Simplified TMC model for the City of Warsaw. Decision tree.

The model illustrated in Fig. 8 reveals that people tend to walk if the walking distance is no longer than 1390m. Otherwise, and assuming they have a driving license, people tend to travel by car when $\frac{d_{\mathrm{CAR}}(j)}{min(d_{\mathrm{PT}}(j))} < 0.47$ i.e. when the travel duration required when relying on the quickest connection by PT

[5] While it is not the objective of this work to describe the process of model development, let us note that some further details on TMC modelling for the City of Warsaw that we rely on in this work can be found in [4].

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

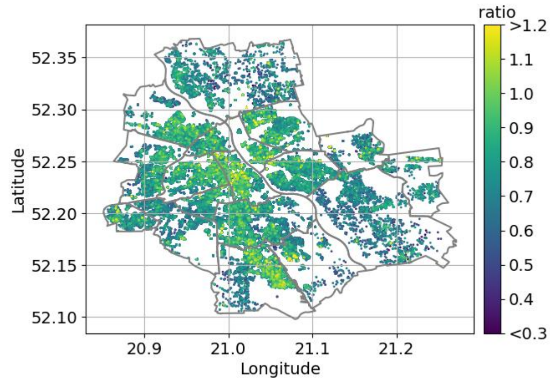


Fig. 9: Travel duration by car vs. public transport. `SYNTH_WAW_INHW` data.

is at least twice as large as the duration of travel by car under free vehicle flow conditions denoted by $d_{\mathrm{CAR}}(j)$. Let us analyse for which journeys this is likely to happen. Fig. 9 shows the spatial distribution of the values of the `minDurationRatioCarToTransit_DIFF` feature. It follows from the figure that while in some (mostly western) parts of the city, areas exist for which travel by PT is even faster than by car, in some parts of the city $\frac{d_{\mathrm{CAR}}(j)}{min(d_{\mathrm{PT}}(j))} < 0.4$ and it may be even three times faster to travel by car (without traffic jams) than by PT. Interestingly, even in such parts as the north part of the City, depending on the exact origin location and trip destination, travellers from these areas are provided with competitive or in rare cases inevitably less satisfactory PT services. This highlights the role of fine-grained spatial analysis of trip features.

4.4 Data needs of the system

Let us note that the main challenge for both the proposed and similar solutions is the need to obtain data needed to estimate the probabilities of exact travel parameters necessary for the generation of the journeys. Without TD matrices and address point data, it is hard to generate realistic coordinates of trip endpoints. We used TD data and address points including the number of house inhabitants for journey generation for our experiments. The number of house inhabitants can be used to estimate the probability of the trips from/to residential buildings. However, this is not the case for commercial buildings. For buildings such as shopping centers estimating at a city scale the number of arriving/departing persons and the origin and destination of their travel can be difficult. Still, journeys for zone combinations can be generated and if zones have moderate area the impact of the problems discussed above is limited.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

5 Conclusions

Travellers in urban areas have to select travel modes for individual trips they make. While for short-distance trips, walking and cycling are most frequently used, sustainable mobility aims at the reduction of private car use. This means inter alia increasing the use of public transport. However, multiple factors influence the decisions of travellers. Furthermore, the overall travel duration depends on walking distances and waiting times such as waiting for transfers.

In this study, we propose a system combining Apache platforms, OpenTripPlanner and transport model data to generate and analyse representative trips in urban areas. The system enables an in-depth understanding of the distribution of trips, their routes and LOS features. Once used for spatial analysis, these data enable the understanding of which areas benefit from short walking distances needed to use PT connections, direct connections and overall PT travel duration comparable to travel duration by car. This provides interesting opportunities for analysing travel mode choices across different city areas. In the future, we plan to further exploit the visualisation of synthetic trip data. An interesting challenge is the use of clustering techniques for these data, though the existence of spatially close points with substantially different feature values confirms the complexity of PT service level patterns.

Acknowledgments

The research leading to these results has received funding from the EEA / Norway Grants 2014-2021 through the National Centre for Research and Development. CoMobility benefits from a 2.05 million€ grant from Iceland, Liechtenstein and Norway through the EEA Grants. The aim of the project is to provide a package of tools and methods for the co-creation of sustainable mobility in urban spaces.

Data made public by the City of Warsaw including schedules and GPS traces from the City Open Data portal (http://api.um.warszawa.pl) acquired and processed in the years 2022-2023 were used to develop public transport schedules. Further details on these data can be found on the Open Data portal.

References

1. Braga, C.K.V., Loureiro, C.F.G., Pereira, R.H.: Evaluating the impact of public transport travel time inaccuracy and variability on socio-spatial inequalities in accessibility. Journal of Transport Geography **109** (5 2023). https://doi.org/10.1016/j.jtrangeo.2023.103590
2. Brunner, H., Hirz, M., Hirschberg, W., Fallast, K.: Evaluation of various means of transport for urban areas. Energy, Sustainability and Society **8**(1), 9 (2018), https://doi.org/10.1186/s13705-018-0149-0
3. Chia, J., Lee, J., Han, H.: How does the location of transfer affect travellers and their choice of travel mode?—A smart spatial analysis approach. Sensors (Switzerland) **20**(16), 1–17 (2020). https://doi.org/10.3390/s20164418

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10

4. Grzenda, M., Luckner, M., Brzozowski, Ł.: Quantifying parking difficulty with transport and prediction models for travel mode choice modelling. In: Mikyška, J., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M. (eds.) Computational Science – ICCS 2023. pp. 505–513. Springer Nature Switzerland, Cham (2023)
5. Hillel, T., Bierlaire, M., Elshafie, M.Z., Jin, Y.: A systematic review of machine learning classification methodologies for modelling passenger mode choice. Journal of Choice Modelling **38**, 100221 (2021). https://doi.org/10.1016/j.jocm.2020.100221, https://doi.org/10.1016/j.jocm.2020.100221
6. Hillel, T., Elshafie, M.Z.E.B., Jin, Y.: Recreating passenger mode choice-sets for transport simulation: A case study of London, UK. Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction **171**, 29–42 (3 2018). https://doi.org/10.1680/jsmic.17.00018
7. Luckner, M., Grzenda, M., Kunicki, R., Legierski, J.: IoT Architecture for Urban Data-Centric Services and Applications. ACM Transactions on Internet Technology **20**(3), 1–30 (2020)
8. Moreno, C., Allam, Z., Chabaud, D., Gall, C., Pratlong, F.: Introducing the "15-minute city": Sustainability, resilience and place identity in future post-pandemic cities. Smart Cities **4**(1), 93–111 (2021)
9. OpenStreetMap contributors: Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org (2017)
10. Redman, L., Friman, M., Gärling, T., Hartig, T.: Quality attributes of public transport that attract car users: A research review. Transport Policy **25**, 119–127 (2013). https://doi.org/10.1016/j.tranpol.2012.11.005
11. Rocha, S.S., Pitombo, C.S., Costa, L.H.M., Marques, S.d.F.: Applying optimization algorithms for spatial estimation of travel demand variables. Transportation Research Interdisciplinary Perspectives **10**(April), 100369 (2021). https://doi.org/10.1016/j.trip.2021.100369, https://doi.org/10.1016/j.trip.2021.100369
12. Sousa Santos, G., Sundvor, I., Vogt, M., Grythe, H., Haug, T.W., Høiskar, B.A., Tarrason, L.: Evaluation of traffic control measures in Oslo region and its effect on current air quality policies in Norway. Transport Policy **99**(August), 251–261 (2020). https://doi.org/10.1016/j.tranpol.2020.08.025
13. Tenkanen, H., Toivonen, T.: Longitudinal spatial dataset on travel times and distances by different travel modes in Helsinki Region. Scientific Data **7**(1) (dec 2020). https://doi.org/10.1038/s41597-020-0413-y
14. Waury, R., Dolog, P., Jensen, C.S., Torp, K.: Analyzing trajectories using a path-based api. In: Proceedings of the 16th International Symposium on Spatial and Temporal Databases. p. 198–201. SSTD '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3340964.3340990
15. Yousefzadeh Barri, E., Farber, S., Jahanshahi, H., Beyazit, E.: Understanding transit ridership in an equity context through a comparison of statistical and machine learning algorithms. Journal of Transport Geography **105**(August), 103482 (2022). https://doi.org/10.1016/j.jtrangeo.2022.103482
16. Zhang, G., Wang, D., Cai, Z., Zeng, J.: Competitiveness of public transit considering travel time reliability: A case study for commuter trips in Hangzhou, China. Journal of Transport Geography **114** (1 2024). https://doi.org/10.1016/j.jtrangeo.2023.103768

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63783-4_10