Inference of Dynamic Origin-Destination Matrices with Trip and Transfer Status from Individual Smart Card Data

Sofia Cerqueira
INESC-ID and Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal
Elisabete Arsenio
LNEC, Department of Transport, Lisbon, Portugal
Rui Henriques
INESC-ID and Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal

ABSTRACT

The provision of seamless public transport supply requires a complete understanding of the real traffic dynamics, comprising origin-to-destination multimodal mobility patterns along the transport network. However, most current solutions are centred on the volumetric analysis of passengers' flows, generally neglecting transfer, walking, and waiting needs, as well as the changes in the mobility patterns with the calendar and user profile. These challenges prevent a comprehensive assessment of the routing and scheduling vulnerabilities of (multimodal) public transport networks.

The research presented in this paper aims at addressing these challenges by proposing a novel approach that extends the inference of dynamic Origin-Destination (OD) matrices to capture vulnerabilities and mobility patterns from individual trip record data. Given specific spatial and temporal criteria, the proposed methodology is able to infer dynamic matrices of the real state of a multimodal transport network, disclosing metrics such as volume, time, distances, transfers and multimodality-related indicators. More specifically, the developed approach offers two major contributions: i) efficient identification of network vulnerabilities pertaining to time and distance spent on transfers and trips; and ii) decomposition of traffic flows in accordance with calendrical rules and user profiles.

The Lisbon's public transport network is selected as the case study for assessing these contributions. The gathered results from smart card validations from the two main public transport operators in the city of Lisbon (bus and metro) reveal vulnerabilities in specific parishes. Also, statistics pertaining to travel time, transfer time, and number transfers reveal disparities between several OD pairs in the city, highlighting differences on accessibility. Research findings are actionable, offering the opportunity for public carriers and municipalities to pursue their efforts towards sustainable mobility.

1. INTRODUCTION

European Urban centres are pursuing sustainable mobility, discouraging individual car-based travel and reinforcing the service quality of public transport. In this context, it is of utmost importance for policy makers to have a clear and comprehensive knowledge of the public transport network (all modes), being able to assess its adequacy, vulnerabilities and unsatisfied mobility needs along time based on passenger dynamics patterns and relevant context information (e.g. socioeconomic). Classic origin-destination (OD) matrices are one of the most used tools by public transport agencies to this end, enabling the analysis of the distribution of passengers' flows along the network. However, this classic method is hampered by multiple obstacles: i) the need to account for ongoing changes in demand and isolate calendar-specific traffic flows (dynamic stance); ii) the need to integrate traffic views from different carriers and modes of transport; iii) the relevance of offering parameterizable spatial resolutions; iv) the importance of providing filters to focus on specific routes and user profiles; and, finally and foremost, v) the need to go beyond classic volumetric views and capture important statistics that can reveal vulnerabilities, such as the distribution of the number of required transfers, as well as the time and distance spent in trips or within transfers throughout the network.

The research presented in this paper addresses the above issues by proposing a novel approach to infer dynamic Origin-Destination (OD) matrices from smart-card validations gathered from (multimodal) public transport networks. More specifically, we discuss three major contributions: i) the estimation of records corresponding to missing stop alighting data; ii) the detection of vulnerabilities on the network pertaining to walking distances and trip durations in a more efficient way, and iii) the decomposition of traffic flows in accordance with calendrical rules and user (passenger) profiles.

Differently from other previous works, these contributions go beyond the analysis of the demand distribution on the network. In fact, these extended OD matrices show the benefits of displaying the network real state according to important statistics, including travel time, transfer time, travel distances, transfer distances, the average number of transfers, and multimodal commutes.

The contributions are validated on the bus-and-metro public transport network in the city of Lisbon. In particular, this work is conducted in the context of the ILU project (Leite, et al. 2020), an innovative project established on advances from artificial intelligence, big data analytics, and urban computing, applied to the integrative analysis and optimization of urban traffic in the Lisbon city.

With the support and validation of the primary public bus operator, CARRIS, an robust and usable tool for the visual analysis of the proposed dynamic OD matrices was further developed. The tool allows several filters to build the OD matrix, including temporal restrictions (time periods, calendrical constraints), spatial granularities (TAZ, parishes, neighbourhood sections, stops), selection of user profiles, trip typologies, amongst other facilities. In the end, it projects the OD matrix onto a heatmatrix, where one of the metrics is highlighted and the remaining metrics are shown in a tooltip that is visible by hovering over the cell. Observing the related literature, to the best of our knowledge, the contributions explained here are unique and encourage a new spatiotemporal perspective of urban traffic.

The paper is organised as follows: section 2 introduces essential concepts pertaining to this multidisciplinary research scope; section 3 identifies the related work on the inference of OD matrices; section 4 introduces the case study; section 5 describes the proposed methodology for the inference of dynamic and multimodal OD matrices; section 6 presents the main research results and implications; finally, major concluding remarks are drawn in section 7.

2. BACKGROUND

2.1 Smart card data and Automatic Fare Collection systems

The monitoring and planning of a public transport system are essential to establish an equilibrium between demand and supply, operation and policy-making (Lu et al. 2020). Commonly, the transport agencies support the planning decisions through the observation of the extracted information from a data-collecting technology. Automated Fare Collection (AFC) systems record passenger entries and/or exits on the network via smart card validation. When a passenger validates the smart card, on the station or in a public vehicle, a record is stored with the timestamp, location, and optional route specifications. In this work, the act of validating a card is called a transaction. AFC systems can be classified as an entry-exit system or a close-system (Mosallanejad et al. 2019, Nassir et al. 2015, Tang et al. 2020). In a close-system, the passenger has to validate the card when both arriving and leaving a station (or vehicle boarding and alighting). A transport network with an entry-only system requires ticket validation only at the boarding. Since the alighting information is not recorded on the entry-only system, the agencies don't know the vehicle load at a given moment and the destination of its passengers, hindering the service planning and management. This is the case of the bus public transport system in the city of Lisbon.

2.2 Trip typology definition

A **trip stage** s: $\theta_{start} \mapsto \theta_{end}$, is a movement of a passenger p without transfers between stop coordinates θ_{start} and θ_{end}, through transport modes (metro, bus, bike, car, among others). The path of a passenger to its final destination is a set S of one or more $(1 \ldots m)$ travel steps $S = \{s_1, s_2, \ldots, s_m\}$. Since some systems are entry-only and, hence, only collect the boarding information, it is necessary for the identification of alighting information to have a complete trip stage record. In the literature, the most studied methodologies are based on rule-based chaining of trip stages (Li et al. 2007, Zhao et al. 2007, Farzin et al. 2008, Li et al. 2011, Nassir et al. 2011, Munizanga et al. 2014, Nunes et al. 2015, Hora et al. 2017, Trépanier et al. 2017, Barry et al. 2019), where the most used rules are the ones enunciated by Barry et al. (2002): i) passengers tend to start their next trip near the exit on the previous trip; ii) the alighting place of the last trip is the same place as they boarding on the first trip of the day. Later, this principle was improved by Trépanier et al. (2007), by suggesting that the last trip is the first boarding place on the day that could be closed and not necessarily the same location. This revised principle is particularly important (or prone to occur) in bus networks because on a given route that has ascending and descending directions, their stops correspond to different locations.

Let a **journey**, j: $\theta_{start} \mapsto \theta_{end}$, to be the movement of a passenger from a origin θ_{start} to a final destination of the passenger's trip θ_{end}, with zero or more transfers through one or more modes of transport. From a set of m trip stages $S = \{s_1, s_2, \ldots, s_m\}$ it is inferred a set of n journeys $J = \{j_1, j_2, \ldots, j_n\}$, were $n <= m$. In the literature, the methodologies to identify the origin and final destination of the passenger's trip are mostly based on the distinction of transfers from an activity (Nassir et al. 2011, Ali et al. 2016). In detail, if the time interval between two trip stages is greater than a certain threshold it indicates that the passenger is doing an activity (work, shopping, home), otherwise, it is a transfer between trip stages. For instance, Alsger et al. (2015) used this methodology to generate origin-destination matrices based on journeys and demonstrated that the transfer time of 15 to 90 minutes had an insignificant impact on the OD matrices.

In commuting trips, the distinction between transfer and activity is simpler to identify due to their periodicity, frequency and the large time discrepancy between the time spent on an activity and on a transfer, such as school-home and vice versa, or work-home and vice versa.

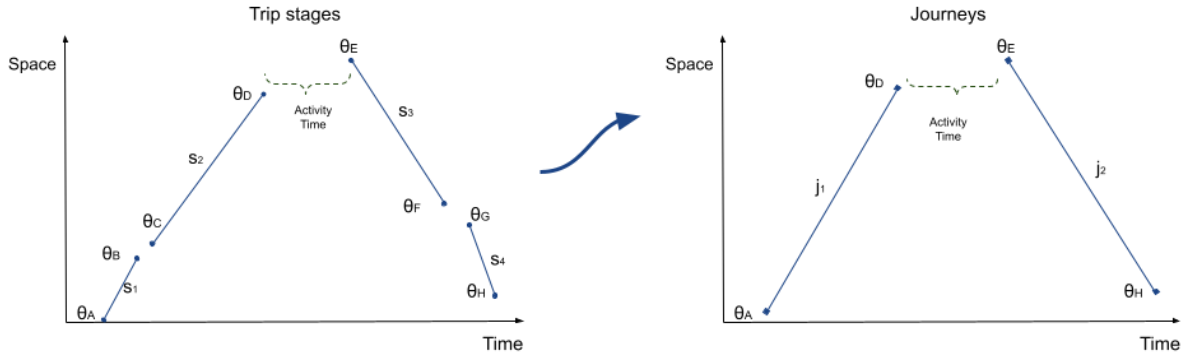

Figure 1: Illustrative trip typology chains for a given passenger during the day. On the left axis it shows the *trip* stage chain. The right side shows the chain of *journeys*, resulting from the identification of transfers and each activity time.

2.3 Origin Destination Matrix

After performing stage trip or journey generation and extraction, we can represent the volumetric distribution of trips, in space, in an origin-destination matrix (Dragu et al. 2019). Each cell of the origin-destination matrix specifies the volume $v_{i,j}$ between an origin i and a destination i. In short, matrices include three modelling features, which are: i) static or dynamic matrix; ii) spatial granularity; and iii) trips typology.

		destination j				
		1	*2*	...	*w*	$O_j = \sum_{i=1}^{n} v_{i,j}$
	1	$v_{1,1}$	$v_{1,2}$	$v_{1,j}$	$v_{1,w}$	O_1
	2	$v_{2,1}$	$v_{2,2}$	$v_{2,j}$	$v_{2,w}$	O_2
origin i	...	$v_{i,1}$	$v_{i,2}$	$v_{i,j}$	$v_{i,w}$	O_i
	z	$v_{z,1}$	$v_{z,2}$	$v_{z,j}$	$v_{z,w}$	O_z
	$D_j == \sum_{i=1}^{n} v_{i,j}$	D_1	D_2	D_j	D_w	$V = \sum_{i,j}^{n\times n} v_{i,j}$

Table 1: Illustrative OD matrix showing the traffic flow volume between boarding entries and *m* alighting entries. The last row and column shows the total volume on a given entry or exit, respectively

Firstly, OD matrices can be classified as either static or dynamic. A static OD matrix considers time-independent flows over the space (Mungthanya et al. 2019). For this typology, methodologies have been developed to capture average flows between OD pairs within a geographic area, in a single matrix, such as gravity models, entropy maximization, information minimization (Yang et al. 2019). However, the advancement of technologies, computational processing and storage resources enabled the inference of dynamic OD matrices. Consequently, dynamic OD matrices become the focus in the study of transportation planning, since it shows more accurately traffic dynamics between zones.

Secondly, in the transport context, the space dimension can be configured for micro and macroscopic analysis. Some studies perform exhaustive and microscopic analysis, observing the flow of trips between pairs of network stops (subway stations, bus stops, bike stations, among others) (Hora et al. 2017). On the other hand, there are matrix studies at the macroscopic level, i.e. between network zones (aggregations of stops), such as TAZ (Mungthanya et al. 2019, Sobral et al. 2021), city parishes, clusters (Luo et al. 2017).
Finally, the content of the matrices can be modelled from trip stages or journeys. Matrices that present the flow in the network through trip stages show the actual passenger volume at all points in the network. While the rendering of journeys-based matrices aims at identifying potential producer and attractor points in the network.

3. RELATED WORK

This section summarizes the related work on the inference of origin-destination matrices and similar contributions in urban traffic visualization. In literature, the inference of origin-destination matrices has a common purpose, which is passenger flow analysis. However, matrices' design diverges in different aspects, that will be herein addressed, in the following order: i) data source ii) temporal and spatial granularities, iii) visualization facilities.
First approaches for the estimation of origin-destination (OD) matrices were based on statistical inference from interviews or/and surveys. However, with the monitoring of individual movements in the network, it has been possible to model dynamic and more accurate matrices of the state of urban traffic through sensory data sources such as phone mobile records (Alexander et al. 2015), global position system trajectories (Mungthanya et al. 2019), and smart card records (Munizaga et al. 2012). In fact, most studies in the scope of public urban transport with AFC systems are dependent on smart card information. For instance, Munizaga et al. (2012) used smart card data from the multimodal public transport system of Chile (metro and bus) to enrich the alighting bus information and apply the bus data to infer OD matrices. Similarly, in 2017, Hora et al. (2017) contributed with an approach that includes smart card data from the transport modes metro, bus and tram. The matrix proposed by Hora et al. (2017) depicts dynamic OD matrices with flow distribution between city Porto zones, where each zone aggregates stops of all transport modes.

The second fundamental point for the design and analysis of OD matrices is spatial granularity. Usually, the explored granularities in literature and transport planning practice correspond to aggregations of stops, such as TAZ, clusters, or zones chosen by the author. According to McCord et al. (2012), stop-to-stop OD matrices make it difficult to explore important pattern flows. Yet, Sobral et al. (2021) states that it is essential to depict OD matrices with several levels of spatio-temporal granularity to encourage use by stakeholders in exploring urban mobility flows. Luo et al. (2017) proposes aggregation of stops through the clustering algorithm K-means. Spatial K-means requires the optimal parametrization of the optimal number of clusters, by maximizing the ratio of average intra-cluster flow to average inter-cluster flow while maintaining the spatial compactness of all clusters. Moreover, Luo et al. indicates clustering aggregation is appropriate for areas with high station density, such as the addressed study case, the city Haaglanden. In terms of computational efficiency, the Luo et al. approach employed in a tool to depict dynamic OD matrices, would be expensive, since it requires clustering training each time it would be queried a different distribution flow. The cost is less when it is assigned stops to each zone before inferring matrices.

Data visualisation plays an indispensable role in the organisation and perception of data. According to Lee et al. (2020), complex data encoded in numbers and text is much more incomprehensible to humans than those visualized in graphics. Indeed, finding useful patterns and information, to the naked eye, in a large set of numerical data, such as target origin matrices, is easier and appealing when the data is expressed in appropriate graphical representations. In the existing literature, we find several options to depict the demand flow between origin-destination, including heat matrices (Sobral et al. 2021), heat maps (Yu et al. 2017, Wood et al. 2010), flow maps (Ibarra et al. 2016), chord diagrams (Wang et al. 2020), and sankey diagrams (Sobral et al. 2019). In the scope of heat matrices, Sobral et al. (2021) proposes a knowledge-assisted visualization tool suitable for each stakeholder Porto. The dashboard allows the visualization of OD matrices with journey flow in Porto's public transport system, with spatial options such as stops, neighbourhood and TAZ spatial granularity. On the other hand, the temporal granularities available are restricted to a daytime windows, such as morning (AM) peaks, afternoon (PM) peaks, and weekends. Each cell of the matrix has a hue color belonging to the scale to indicate the volume between the OD pair and the flow details are shown after hovering the mouse over a cell. Interestingly, if the matrix has a coarse granularity e.g. neighbourhood, clicking on a cell drills down the spatial perspective by one level, e.g. stops. Similarly, the application developed within the scope of our research was tailored to the Lisbon bus network, however, it shows statistics beyond the volume of passengers distribution. Further, the temporal granularity of the heat matrices is not fixed in window times. The tool allows other queries such as views of profile users types, boarding and alighting restrictions (routes and stops), typology of trips (trip stage and journeys), and more granularities (parishes). Likewise, to Sobral et al. approach, the hovering act displays all information regarding the available metrics.
In contrast with heat matrices, heatmaps offer the complementary advantage of associating passenger demand with geographical locations on the map. However, the display an arbitrarily high number OD pairs in a geographical area can deteriorate usability, and generally inflows and outflows need to be represented in two separate heatmaps. For instance, Yu et al. (2017) details bus travel demand patterns using the heat maps to identify entries and exits, in the Guangzhou bus network, China. Flow patterns were identified at different temporal granularities, including periods of the day, days of the week, weekdays, weekends and vacation periods. Wood et al. (2010) counteracts the restriction of representing OD flows in heatmaps by proposing a new visualisation approach. The developed technique preserves the spatial layout of all origin and destination locations by constructing a gridded two-level spatial treemap. The result is a set of vectors projected on each geographic area.

4. LISBON CITY AS THE STUDY CASE

The smart card transactions used in the scope of this research were made available from the main Lisbon public transport operators, including the main bus operator, CARRIS, and the subway operator, METRO.

A residual number of transactions, with a lack of passenger identifier, boarding timestamp and boarding location, were removed from the original dataset. Briefly, the bus and subway datasets have approximately 11 million and 31 million trip records, respectively, for October 2019. Dataset features with smart card transactions from bus and subway networks are described in Tables 2 and 3, respectively.

Dataset column	Description
Card ID	Card identifier.
Boarding Timestamp	Date and time registered at card validation, aboard the bus vehicle.
Route	Identifier of the route on which the bus is operating.
Direction	Direction of the route (ascending, descending, circular).
Variant	Some routes have several variations covering subparts of the original stop sequence.
Stop code boarding	Code that identifies the boarding bus stop.
Stop name	Name of the boarding stop
Stop sequence boarding	Sequence number of a stop, on a given route.
Card type	Fare code related to the card (ex: Sub18/Sub23 is the student card).

Table 2: Dataset columns from bus data collection.

Dataset column	Description
Card ID	Card identifier.
Boarding Timestamp	Date and time stored at the boarding station.
Boarding Station	Station code stored at the validation on the boarding.
Alighting Timestamp	Date and time registered alighting station.
Alighting station	Station code registered validation on the alighting.

Table 3: Dataset columns from subway data collection.

5. SOLUTION

The proposed methodology for inference of dynamic OD matrices with trip and transfer status is composed by three major steps: i) the completion of trip record data for entry-only systems (section 5.1); ii) the identification of journeys from individual trips (section 5.2); and iii) the estimation and visualization of statics along end-to-end traffic flows in accordance with the selected spatiotemporal criteria and optional filters (section 5.3). Section 5.4 further introduces the proposed tool.

5.1. Alighting estimation of a stage trip

To generate dynamic origin-destination matrices we need complete information of each passenger's trip stages. Therefore, in this subsection, we briefly explain the algorithm for alighting stop and timestamp estimation for transactions collected from the bus entry-only system. The developed algorithm chains the bus and subway transactions to trace the passenger's path, in the bus and metro network. Subsequently, to determine the location of unknown exit bus stops, the model follows the principles described by Barry et al. (2002), explained in section 2.

Figure 2 provides a flowchart illustrating the steps for processing metro and bus transactions of a given passenger (Barry et al. 2002). The model can be parameterized to receive transactions from a suitable time window. Indeed, a 24 hour period was chosen, starting from one day at 03:59:59 to 04:00:00 the next day. Then, the algorithm collects the transactions from a parameterized period ordered by passenger identification and chronologically.

The assessment of candidate alighting stops depends on a suitable distance equation. For our problem, the haversine distance is used, since it expresses more accurately the walking distance and is widely used in recent studies (Hora et al. 2017, Assemi et al. 2020).

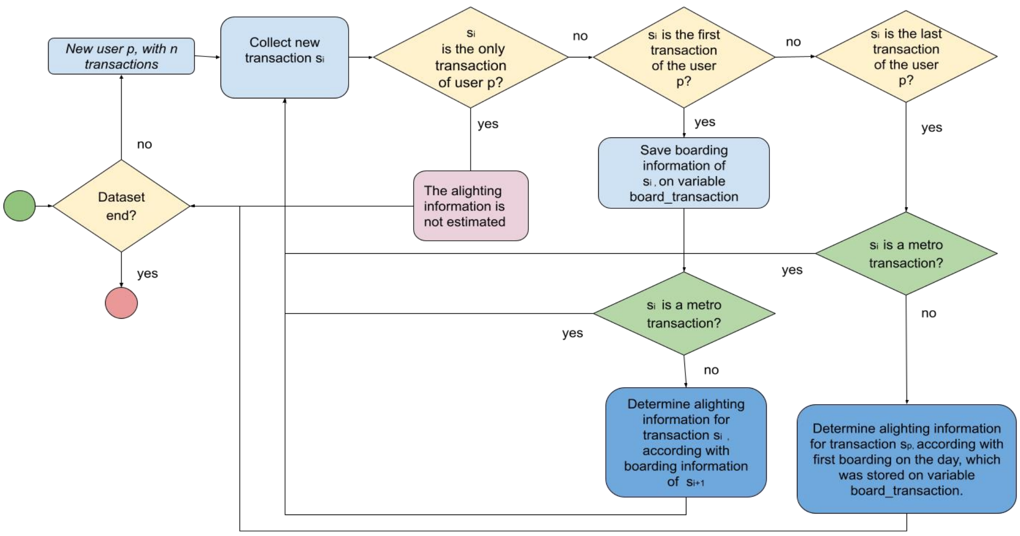

Figure 2: Model flowchart for alighting estimation of a stage trip.

The candidate stop θ that minimizes transfer distance, along a given route, is chosen. However, if the calculated transfer distance exceeds the threshold (parameterizable on the model), the θ is not valid and the transaction remains without alighting information. We consider that the maximum transfer distance should not exceed the 1000 meters threshold (Munizaga et al. 2012). In the end, the model produces a dataset with trip stages, including the following columns shown in Table 4, along with other statistical features.

In the related surveys, the developed approaches are limited to keep boarding and alighting information. In our enhanced solution, we additionally calculate and store relevant statistics, such as travel time, trust level, travel time, including transfer distance (as shown in the Table 4) associated with each trip stage. With this method, the computational effort generated by the model is compensated by the creation of low cost and efficient dynamic matrices based on calculated a priori statistics. On the other hand, storing these statistics offers freedom for further exploratory analysis beyond the classic OD matrices.

5.2 Boarding and alighting stops estimation for a journey

The model proposed in this section aims to generate journeys, whose origin and destination are respectively the beginning of the trip and the final destination (trip purpose) of the passenger. Some common examples that illustrate this typology of journeys are routine or functional trips, such as home-work or home-school commuting.

Journeys are derived from a set of tip stages made during the day. A journey ends when the passenger alights to perform an activity. In the proposed algorithm, the activity is identifiable through the time spent between trip stages. That is, if the time spent between trip stages is greater than the defined threshold, the passenger is considered to be performing an activity; otherwise, it is considered to be a transfer between public transport vehicles. The threshold defined for our case study is 90 minutes (Alsger et al. 2015).

Route code	Direction	Stop code boarding	Datetime boarding	Stop code alighting	Datetime alighting	Transfer Distance (meters)
706	DESC	510	2019-10-01 08:24:04	6608	2019-10-01 08:25:39	0
774	DESC	6608	2019-10-01 08:38:48	1211	2019-10-01 08:41:36	29.5
774	ASC	1216	2019-10-01 17:02:05	6611	2019-10-01 17:07:29	108
706	ASC	6906	2019-10-01 17:40:36	511	2019-10-01 17:42:00	42.4

Table 4: Data outcome from alighting estimation of a stage trip.

Boarding route code	Boarding Direction	Stop code boarding	Datetime boarding	Alighting route code	Alighting Direction	Stop code alighting	Datetime alighting	Transfer time (min)
706	DESC	510	2019-10-01 08:24:04	774	DESC	6608	2019-10-01 08:41:36	789
774	ASC	1216	2019-10-01 17:02:05	706	ASC	511	2019-10-01 17:42:00	1987

Table 5: Data outcome from boarding and alighting estimation of a journey.

Tables 4 and 5 show a real example of the estimation of journeys. The algorithm receives, as input, a dataset with stage trips of a passenger, sorted by boarding date, as shown in Table 4. Finally, the algorithm estimates two journeys, as shown in Table 5. In the afternoon, the passenger makes a return trip, transferring between the same routes, but in the upward direction.

Similarly to the trip stage estimation model, this solution stores other feature statistics associated with each journey, in addition to those indicated in Table 4. In addition to the trip time and distance statistics, the model also calculates the number of transfers made by the passenger during the journey, the time and the total distances spent on the transfers. This new approach allows efficient resource allocation for generating journey-based matrices, dependent on statistical metrics.

5.3 Inference of Dynamic OD matrices

One of the main contributions of this research resides on the inference and visualization steps. The proposed solution aims to overcome traditional presentations centred on the distribution of volume in the network. Considering a given spatial resolution and temporal constraints (time interval and calendrical restrictions), we generate dynamic matrices able to comprehensively describe the real state of the network, through metrics such as volume, time, distances, transfers, and multimodality indicators. This functional multiplicity of matrices allows a detailed and precise identification of vulnerabilities and mobility disparities within the city.

The OD matrices are modelled from three components which are. i) the data source; ii) optional filters including temporal, spatial and user profile restriction; and iii) the selection of one of the possible statistics as primary organization criterion. The next paragraphs explain each of these modelling fields.

Firstly, the approach allows the formation of dynamic matrices based on one of the trip typologies which are either trip stage or journey. Secondly, the matrices can be modelled according to time, space and passenger typology, by parameterizing: a) a desirable time window restricting dates and times; b) the target weekdays (weekend, working days, one or more days of the week); c) the target user profiles; d) the desirable entry and exit routes and stops (by default all the transport network is considered). This last filter is only available for matrices filled with journeys, since the alighting route may not be the same as the boarding one. Third, one of the following metrics must be chosen to guide the organization of the matrix (the highlighted metric) and the remaining is displayed by hovering in each cell (OD pair):

[a] Volume: total trips'volume;
[b] Average volume per day: average daily travel volume;
[c] Number of transfers: average number of transfers between origin and destination;
[d] Transfers volume: total number of transfers made between origin and destination;
[g] Transfer distance: average walking distance spent by passengers to transfer;
[h] Transfer Time: average time spent by passengers to make a transfer;
[d] Travel distance: average distance travelled during a trip;
[d] Travel Time: average time spent during a trip;
[d] Trust level: the percentage that determines the trust of the data used to report cell information. This confidence is calculated through the walking distance in the transfer (200 meters corresponds to 100% and 1000 meters corresponds to 1% reliability).

Principles for the efficient composition of pivot tables are pursued to aggregate information from individual journeys. In this context, multiple estimators are considered per statistic and OD pair, including median, mean, and standard deviation.

The visual representation of the matrices is given through interactive heat matrices with usable zooming and selection facilities. The hue of each cell changes according to a scale that varies between the minimum and maximum value of the highlighted metric. As mentioned before, only one of the aforementioned metrics is chosen to tone the matrix and the rest are coupled and displayed through tooltips.

At the top of the matrix and on the left side are bar charts that summarize information about the total boardings and alightings, respectively. If the highlighted metric in the OD pairs is volume, the bar charts show the sum of the passenger volumes presented in the rows and columns to indicate the total volume of boardings and alightings, respectively. If the highlighted metric is related to time or distance, the average value is weighted according to the cell's volume presented in the row or column.

5.4 Tool for OD inference with trip and transfer status

With the support and validation of the major public bus operator in the city of Lisbon, CARRIS, a robust tool was developed for the guided parameterization and usable visualization of the proposed dynamic OD matrices. The tool allows the specification of several filters, including temporal filters (time windowing, calendar selections), spatial granularities (TAZ, parishes, neighbourhood sections, stops), user profile filters, trip typologies, amongst others. Figure 3 provides a snapshot of the parameterization board. The visualization of OD matrices satisfies strict usability requirements, incorporating zooming, navigation, and exportation facilities. Both heatmatrix and heatmap visualizations are supported, as well as statistical reports for summarization and background checks.

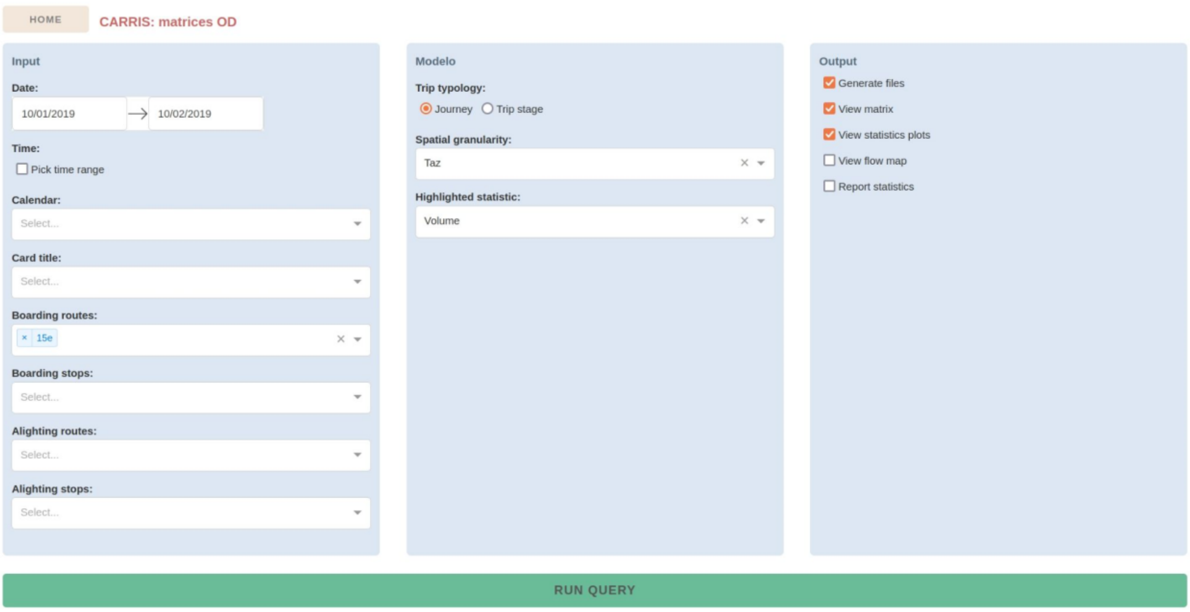

Figure 3: Dashboard responsible for the modelling and parameterization of OD matrices.

6. RESULTS

This section validates the relevance of the proposed contributions using the Lisbon's public transport network as the study case. We want to assess whether the cross-cutting dynamic views over the available statistics can reveal novel knowledge and guide the public transport service. In particular, we explore several scenarios, including age groups distribution, intra-city disparities, and regions connectivity. Since CARRIS, the target bus operator, has an entry-only system, the inferred alighting stops must be validated to verify if the outcomes from models discussed in the sections 5.1 and 5.2 satisfy the principles mentioned in section 2. Following, we show the contributions of this research with illustrative analysis of OD matrices for October 2019.

6.1 Sensitive analysis of trip stages and journeys

Figure 4 shows the distribution of transactions whose alighting stop labels were successfully estimated. The blue bar, whose percentage is 11.6%, indicates the percentage of transactions with no estimated alighting stop. These transactions are unlinked with other transactions, therefore alighting information remains unknown. In the orange bar, we observe that 11.1% of the transactions were chained with other transactions, but the estimated output was not valid, since the inferred alighting bus stop is at a distance above 1000 meters from the boarding of the subsequent transaction. The last bar, green, indicates the success percentage: 77.3% of the input transactions were assigned with a valid alighting stop. Thus, these transactions compose the data source for modelling matrices based on trip stages.

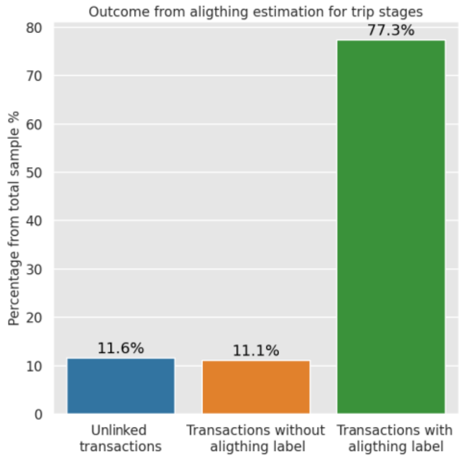

Figure 4: Distribution of bus transactions with and without estimated alighting stop. Transactions in the blue bar are isolated trips, while transaction in the orange bar correspond to estimations without statistical significance due to considerably highly walking distances and waiting times in transfers.

Figure 5 shows the percentage of trip stages (transactions with alighting stop estimated) between a range of the walking distance spent after alighting at the estimated stop. The results are in agreement with the principle that passengers tend to walk as short distances as possible, in transfers and after arriving at their destination. The first bar of the chart does not correspond to an interval because it is intended to highlight that 14.4% destination stop of trip stages is the same as the boarding of the subsequent trip stage. Another conclusion that corroborates the mentioned principle is the distribution of the percentage of trip stages that decreases with the increase of the walking distance interval. Observing the accumulated percentage, we verify that 91.30% of the trip stages, the walking distance on transfers is less than 500 meters, and the remaining percentage is residual and distributed in the remaining intervals.

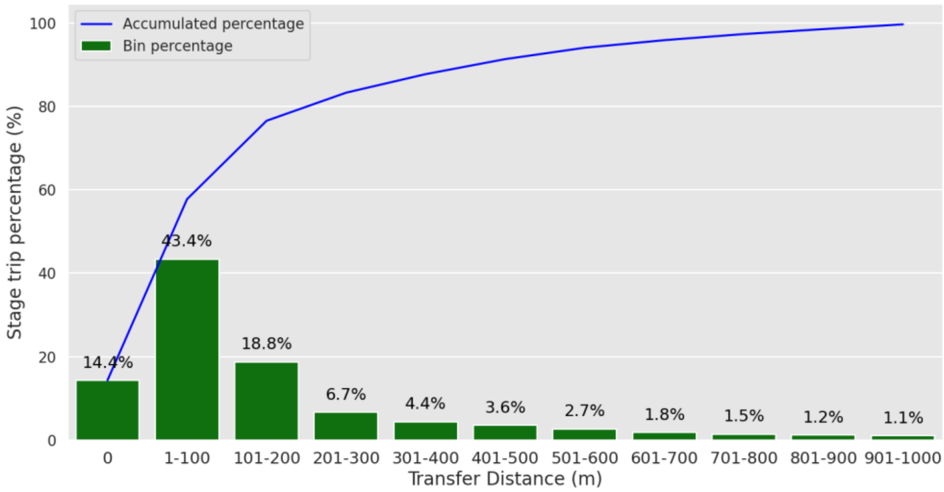

Figure 5: Cumulative and bin percentage of trip stages with a respective transfer distance interval.

Figure 6 describes the distribution of journeys according to the number of transfers. In short, 72.5% of the journeys have no transfers, 21.8% have one transfer and the remaining percentage is residually distributed in a number higher or equal to two transfers. These results corroborate the assumption that passengers prefer to walk the least as possible and consequently achieve the final destination without transfers.

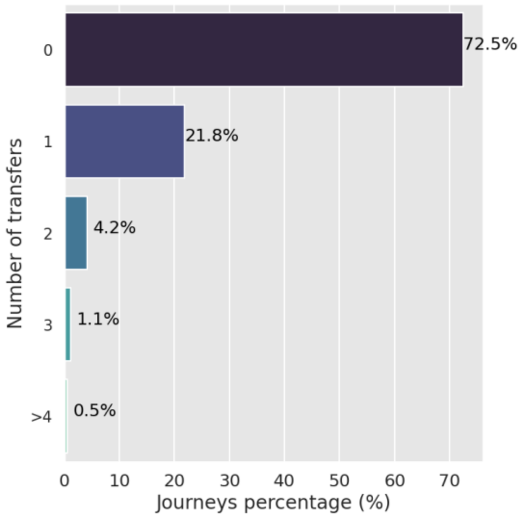

Figure 6: Distribution of journeys according to the number of transfers.

6.2 Dynamic Origin Destination matrix analysis

In this section, we present the results from the exploration of dynamics OD matrices, that goes beyond the volumetric distribution analysis. Our proposal aims to express a complete and detailed understanding of the reality of dynamic traffic and patterns in a city, through several variables.

The figures are directly taken from the produced output of the application used to visualise OD matrices in different conditions. In common, matrices are based on journeys and parishes of Lisbon are selected as the default spatial granularity. Parishes are the geographical (administrative) divisions of the city of Lisbon; these allows displaying fewer rows and columns in the matrices, yielding a simpler and more understandable visualization. The matrix rows correspond to boarding parishes, the columns are the alighting parishes.

6.2.1 Volumetric OD matrix for user profiles

Studies show policies and restructuring of public transport services targeting age groups such as the young and elderly can offer freedom and independence in their mobility, and implicitly generate a positive impact on their lifestyles (Reinhard et al. 2018, Papa et al. 2018, Levin et al. 2019). Therefore, Figure 8 motivates the potentiality of the OD matrices to assess the traffic dynamics of specific age groups, wherein this case is the elderly group.

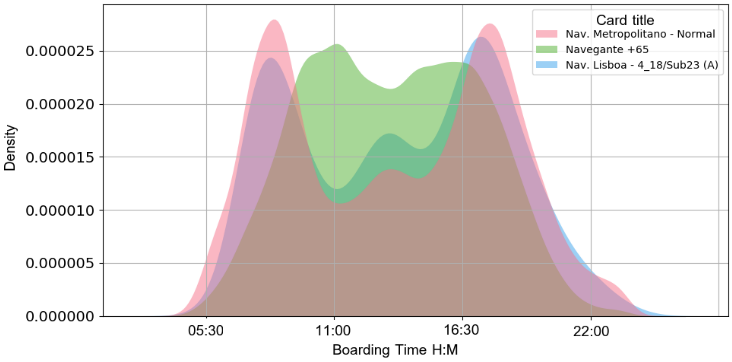

Figure 7: Distribution of boarding on the network for one day per profile user, smoothed by a kernel density function. The blue, green and pink areas show the boarding distribution of younger, elder and adult profiles.

Furthermore, the results of Figure 7 show that the distribution of demand of the elderly age group is slightly different from other age groups. According to Figure 7 elderly individuals (green wave) concentrate their travel from 10 am to 11 am and from 4 pm to 5 pm. These results are in line with evidence from Szeto et al. (2017) study, where it indicates that older people choose 10 am to 11 am to avoid overcrowded public transport (as shown in our results by peaks in pink and blue wave representing the density of entries in the network for adult and young age groups, respectively).

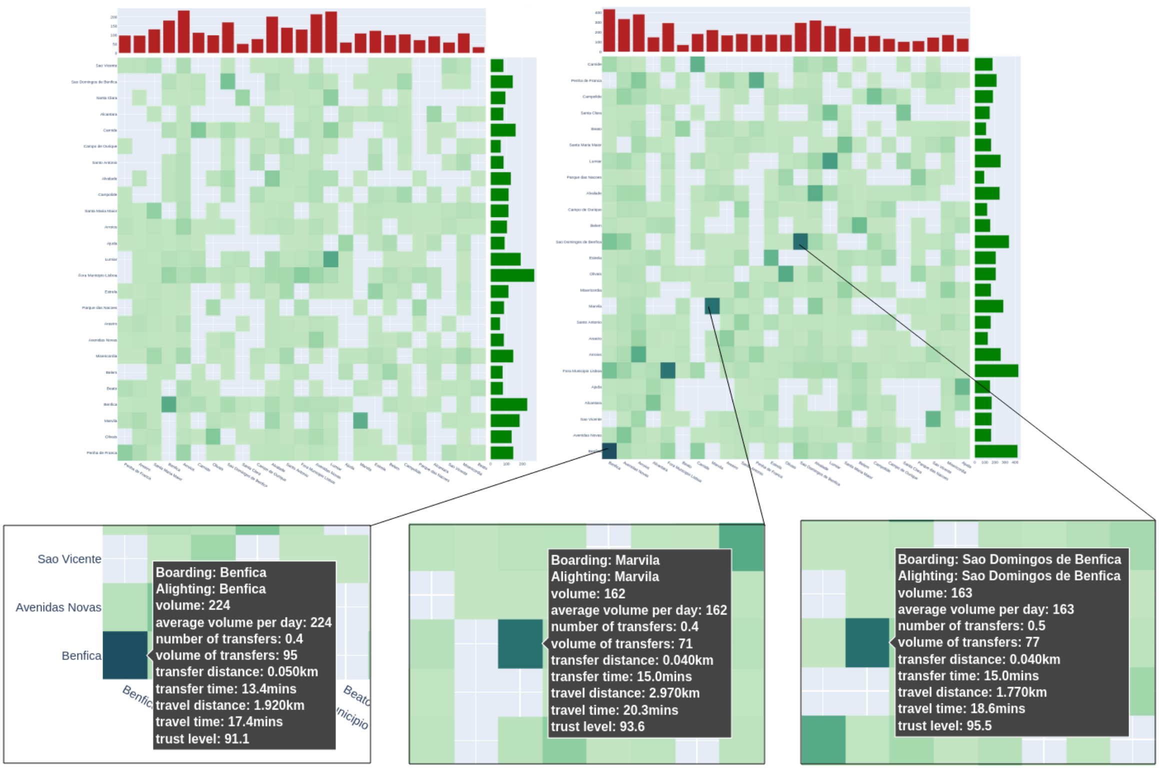

Figure 8: Matrices OD showing flow distribution of the elderly age group by two time windows, between 7 am and 9 am and 9 am and 11 am. The matrices' scale range from 0 to 200 and the elderly age group is represented by these two card titles - Navigator+65 and Urban Navigator 3rd age.

The statistical highlighted metric (which determines the hue of the cells) in the matrices at the Figure 8 is the volume flow. Additionally, at the top and on the right side, the bar charts indicate the total volume of boarding and alighting on the network, respectively. The left matrix corresponds to the period between 7 am and 9 am and on the right side we see the matrix from the period between 9 am and 11 am of October 2nd (Wednesday). The cell shading between the matrices of Figure 8 reaffirms the fact that during the period between 9 am and 11 am there is higher traffic of elderly passengers. It is also easily observed that the largest volume is between the following OD pairs: Benfica-Benfica; Marvila-Marvila; São Domingos de Benfica-São Domingos de Benfica, i.e., there's an higher internal mobility dynamics (within parishes) than between different parishes. This fact matches the information of the number of elderly residents per parish described in the 2011 decennial census, where the parishes with the highest number of elderly residents are São Domingos de Benfica, Benfica and Marvila.

In both matrices, the cells (parish-parish OD pair) with the highest volume are those representing traffic within the same parish. This pattern indicates that older people prefer to travel shorter distances and within the same parish. Furthermore, the results of this figure corroborate with evidence of the study by Wong et al. (2018), which states that shortening the walking and waiting times and improving seat availability can improve the probability of the elderly making a trip. Therefore, public transport can be a means to promote more active lifestyles for the elderly. Assessing the detailed information of some cells, by hovering over, other valuable information can be revealed. For instance: i) the average transfer distance is low, ranging between 4 to 50 metres; ii) the trip distance and travel time are relatively short, 1.7 to 2.9km and 13.4 to 15 minutes, respectively; iii) the average number of transfers ranges between 0.4 to 0.5, and the total number trips with transfers ranges between 71 to 95. The latter metric indicates that around 42% to 47% of trips require a transfer. We conclude that the connectivity between locations inside of the parish could be improved to benefit the elderly group, by dedicating mobility services, such as neighbourhood routes.

6.2.2 OD matrix based on average number of transfers

Figure 9, shows the average number of transfers between origin-destination pairs in a 24 hour period, on October 9. We zoomed two cells that show vulnerability in OD pairs. According to Figure 9, the cell with the highest mean number of transfers is the pair Penha de França-Santa Clara (entry-exit). The detailed information, on the cell, shows that it takes on average 2 transfers to move between Penha de França and Santa Clara. Moreover, the average transfer time (124 min) and the average trip time are extremely high (66.7min). Despite the low number of trips, the results show that the connectivity between these parishes is disparate regarding the rest of the network.

The second cell zoomed at the bottom shows the detailed information about Olivais to Olivais (traffic within the parish). At first sight, the highlighted indicator (average number of transfers) seems low, corresponding to 0.4 (close to 0 transfers). However, if we observe the exact value of the volume of trips and the volume of transfers, 1926 and 764, respectively, we verify that approximately 40% of the trips required 1 or more transfers within the same parish. These findings reveal that the public transport network can be improved within the parish of Olivais to enable better user-place connectivity. Actually, according to Suman et al. (2019), decreasing transfers is the key to encouraging bus transport use and further states that improving connectivity saves users travel time.

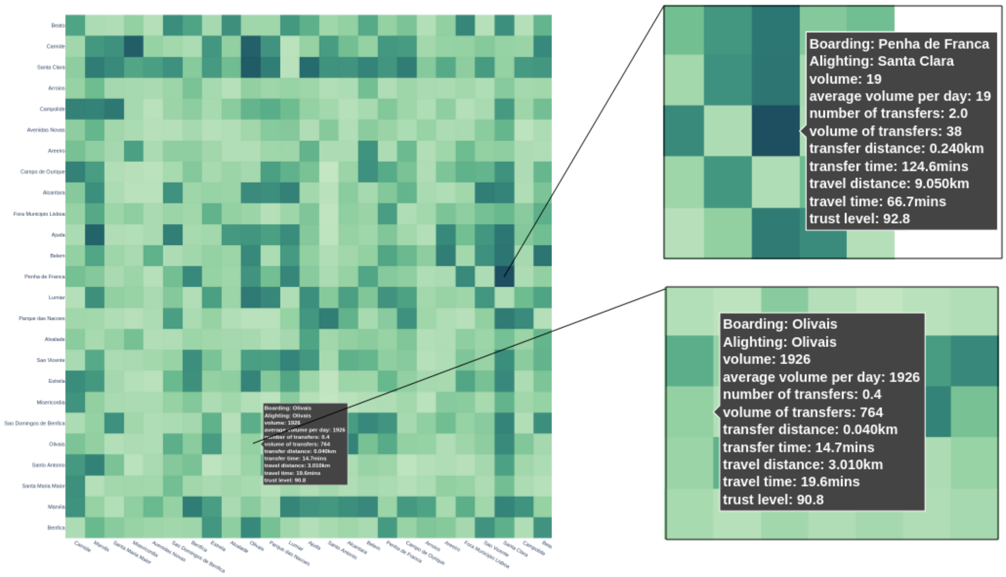

Figure 9: OD matrix showing the average value of transfers required to transfer between an origin and destination; the scale ranges from 0 to 2.

6.2.3 OD matrix analysis on average travel time metric

Figure 10 (a) presents patterns regarding the average travel time between each origin and destination, within a 24 hour period on 2 October. Moreover, the top bars are able to show the average time required to reach a given destination from any point on the network and left bars to show the average time from a given origin to any point in the network, respectively.

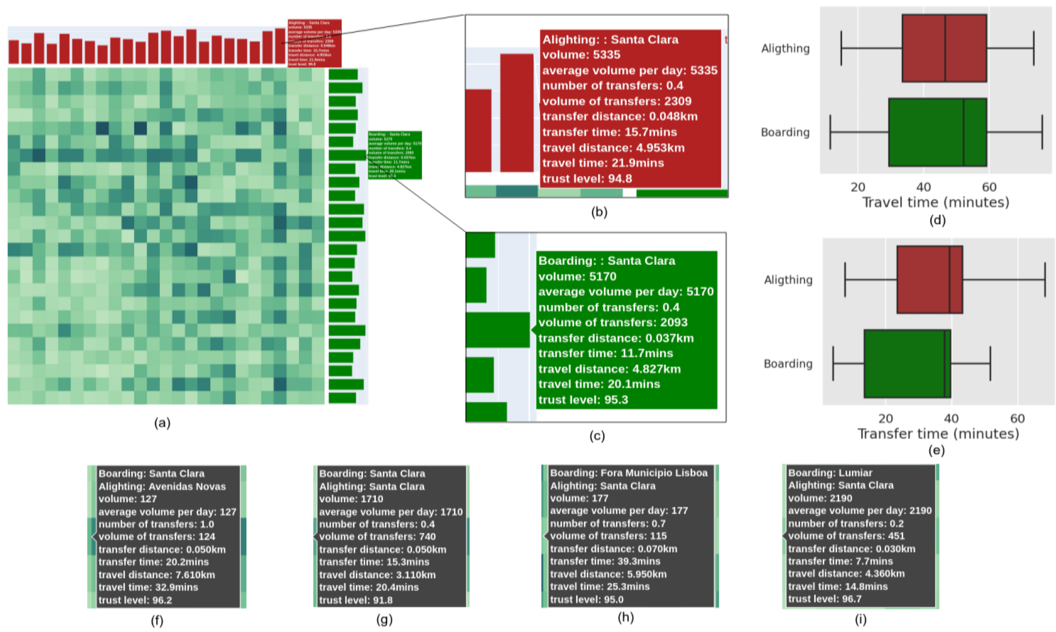

Figure 10: OD matrix showing the average travel time between a origin and destination.

This indicator and the set of visuals identify different areas on the network whose accessibility in terms of average travel time is higher or lower. For instance, the bar charts in Figure 10 (a) show evidence that the parish with the longest average arrival and departure times is Santa Clara, 21.9 minutes and 20.1 minutes. This average time at departure seems reasonable, however, it is weighted with the volume of each cell, and therefore OD pair cells with higher volume and low average time may be hiding critical cases of OD pairs with the lower flow but with higher travel times. In fact, the detailed information on the cells at the Figure 10 (f), (g) and (h), (i) reinforce the evidence of inaccessibility to entry and exit on the Santa Clara parish. The zoomed cells (f) and (g), on Figure 10, show the OD pairs Santa Clara-Avenidas Novas and Santa Clara-Santa Clara with the respective volume 127 and 1710, and the average travel times are 32 and 20 minutes, which seems moderate travel time. However, if we add the average transfer time to the travel time, the total time spent on the trips for each pair OD is 53 and 35 minutes for 7.6 km and 3.1 km average trip distance, respectively. These statistical indicators show strong evidence that Santa Clara must be a target for new route modelling. The same scenario with volume and temporal difference happens as well with the zoomed cell (h) and (i) where the alighting parish is Santa Clara.

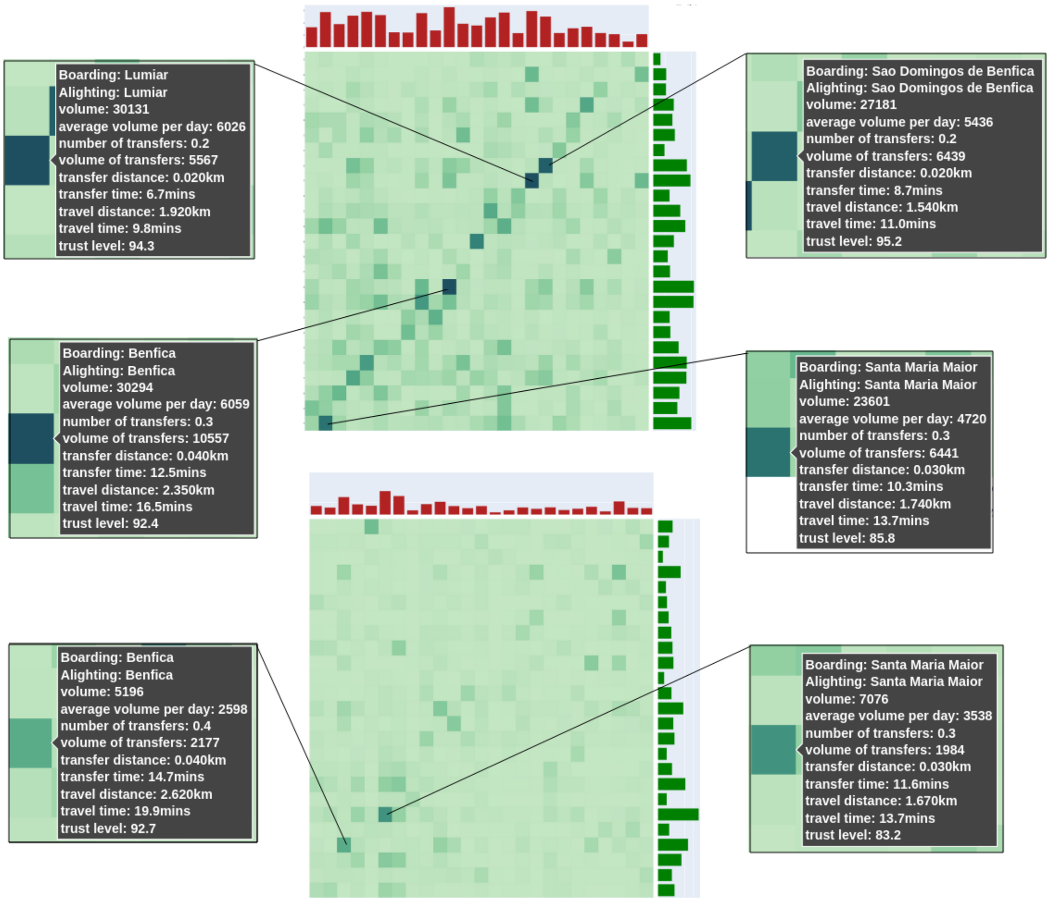

Figure 11: Matrices comparing average volume by day between different days intervals , which are working days (top matrix) and weekend days.

6.2.4 OD matrix for calendrical periods

The highlighted metric in the OD matrices presented in Figure 11 is the average daily volume, in different contexts. The top matrix corresponds to the weekday period (7 to 11 October) and the matrix below represents a weekend (12 to 13 October). The scale of the matrices ranges from 0 to 6000, and the scale of the bar charts ranges from 0 to 17000. In both matrices, we show the cells with the highest daily volume number, zoomed on the sides of Figure 11.

As expected, the average daily volume on weekdays is higher than at the weekend. And in both matrices, the OD pairs with higher volume correspond to traffic within parishes with higher resident density. The OD pair Santa Maria Maior - Santa Maria Maior, in the period of working days, is the fourth pair with the highest average daily volume. However, the weekend becomes the parish with higher internal traffic and more daily entries. This situation may be explained by the touristic flows in Lisbon since this parish corresponds to the Lisbon’s historical centre.

7. CONCLUDING REMARKS

The reported research offers a new approach for the analysis of passengers’ flow behaviour and inference of dynamic OD matrices. We propose alighting stop inference models over the passengers’ paths in the absence and presence of multimodal views, offering the possibility to parameterize maximum walking distances and waiting times on route transfers, extending classical assumptions, and further addressing statistical indicators.

Furthermore, the proposed approach for inferring OD matrices yields four unique contributions. First, we allow inference to consider multimodal commuting patterns, detecting individual trips undertaken along with different operators. Second, we support dynamic matrices’ OD inference along with parameterizable time intervals and calendrical rules, and further support the decomposition of traffic flows according to the user profile. Third, we allow parameterization of the desirable spatial granularity and visualization preferences. Fourth, our solution efficiently computes several statistics that support OD matrix analysis, helping with the detection of vulnerabilities throughout the transport network. More specifically, statistical indicators related to travellers’ functional mobility needs (commuters for working purposes, etc.), walking distances and trip durations are supported. The inferred dynamic OD matrices are the outcome of a developed software with strict guarantees of usability.

Results from the case study using data gathered from the two main public transport operators in the city of Lisbon (Bus and Metro) show that 77.3% of alighting stops can be estimated with a high confidence degree from bus smart-card data. Since the analysis of patterns showed that nearly 27,5% of the journeys within Lisbon's transportation network require one or more transfers, the inferred OD matrices allowed the identification of vulnerabilities in the network, offering the bus public operators in Lisbon new knowledge and a means to better understand dynamics and validate OD assumptions.

The dynamic OD matrices explored within the scope of this investigation showed relevant patterns, including evidence of the greater predominance of flows within parishes, by the elderly; factors such as travel time, transfer time and transfer show that there are significant intra-city disparities, with Santa Clara being one of the parishes with significant vulnerabilities, regarding connectivity and accessibility. Research findings are actionable, offering the opportunity for carriers and municipalities to pursue their efforts towards sustainable mobility.

Acknowledgments
The authors thank CARRIS, METRO and *Câmara Municipal de Lisboa* (*Gabinete de Mobilidade* and *Centro de Operações Integrado*) for the provision of data, support and valuable feedback. This work was further supported by national funds through *Fundação para a Ciência e Tecnologia* (FCT) under project ILU (DSAIPA/DS/0111/2018) and INESC-ID pluriannual (UIDB/50021/2020).

BIBLIOGRAPHY

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

Ali, A., Kim, J., and Lee, S. (2016). Travel behavior analysis using smart card data.KSCE Journal of Civil Engineering, 20(4):1532–1539.

Alsger, A. A., Mesbah, M., Ferreira, L., and Safi, H. (2015). Use of smart card fare data to estimate public transport origin–destination matrix.Transportation Research Record, 2535(1):88–96.

Barry, J. J., Freimer, R., and Slavin, H. (2009). Use of entry-only automatic fare collection data to estimate linked transit trips in new york city. Transportation research record, 2112(1):53–61.

Barry, J. J., Newhouser, R., Rahbee, A., and Sayeda, S. (2002). Origin and destination estimation in new york city with automated fare system data .Transportation Research Record, 1817(1):183–187.

Dragu, V. and Roman, E. A. (2019). The origin–destination matrix development. InMATEC Web of Conferences, volume 290, page 06010.EDP Sciences.

Farzin, J. M. (2008). Constructing an automated bus origin–destination matrix using farecard and global positioning system data in sao paulo,brazil.Transportation research record, 2072(1):30–37.

Hora, J., Dias, T. G., Camanho, A., and Sobral, T. (2017). Estimation of origin-destination matrices under automatic fare collection: the case study of porto transportation system.Transportation Research Procedia, 27:664–671.

Ibarra, J. C., Triana, J. A., and Hern´andez, J. T. (2016). Visualization of origin-destination matrices using a connection barchart and coordinated maps. In2016 IEEE Pacific Visualization Symposium (PacificVis), pages 189–193. IEEE.

Lee, B., Choe, E. K., Isenberg, P., Marriott, K., and Stasko, J. (2020). Reaching broader audiences with data visualization.IEEE ComputerGraphics and Applications, 40(2):82–90.

Leite, I., Finamore, A., and Henriques, R. (2020). Context-sensitive modeling of public transport data.

Levin, L. (2019). How may public transport influence the practice of everyday life among younger and older people and how may their practices influence public transport?Social Sciences, 8(3):96.

Li, D., Lin, Y., Zhao, X., Song, H., and Zou, N. (2011). Estimating a transit passenger trip origin-destination matrix using automatic fare collection system. InInternational Conference on Database Systems for Advanced Applications, pages 502–513. Springer.

Lu, K., Liu, J., Zhou, X., and Han, B. (2020). A review of big data applications in urban transit systems.IEEE Transactions on IntelligentTransportation Systems, 22(5):2535–2552.

Luo, D., Cats, O., and van Lint, H. (2017). Constructing transit origin–destination matrices with spatial clustering.Transportation ResearchRecord, 2652(1):39–49.

McCord, M. R., Mishalani, R. G., and Hu, X. (2012). Grouping of bus stops for aggregation of route-level passenger origin–destination flow matrices. Transportation research record, 2277(1):38–48.

MOSALLANEJAD, M., SOMENAHALLI, S., VIJ, A., and MILLS, D. (2019). An approach to distinguish destination from the alightingstop based on fare data.Journal of the Eastern Asia Society for Transportation Studies, 13:1348–1360.

Mungthanya, W., Phithakkitnukoon, S., Demissie, M. G., Kattan, L., Veloso, M., Bento, C., and Ratti, C. (2019). Constructing time-dependent origin-destination matrices with adaptive zoning scheme and measuring their similarities with taxi trajectory data.IEEE Access, 7:77723–77737.

Munizaga, M., Devillaine, F., Navarrete, C., and Silva, D. (2014). Validating travel behavior estimated from smartcard data.TransportationResearch Part C: Emerging Technologies, 44:70–79.

Munizaga, M. A. and Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smart card data from santiago, chile.Transportation Research Part C: Emerging Technologies, 24:9–18.

Nassir, N., Hickman, M., and Ma, Z.-L. (2015). Activity detection and transfer identification for public transit fare card data.Transportation,42(4):683–705.

Nassir, N., Khani, A., Lee, S. G., Noh, H., and Hickman, M. (2011). Transit stop-level origin–destination estimation through use of transit schedule and automated data collection system.Transportation research record, 2263(1):140–150.

Nunes, A. A., Dias, T. G., and e Cunha, J. F. (2015). Passenger journey destination estimation from automated fare collection system data using spatial validation.IEEE transactions on intelligent transportation systems, 17(1):133–142.

Papa, E., Carpentieri, G., and Guida, C. (2018). Measuring walking accessibility to public transport for the elderly: the case of naples.Journal of Land Use, Mobility and Environment, pages 105–116.

Reinhard, E., Courtin, E., Van Lenthe, F. J., and Avendano, M. (2018). Public transport policy, social engagement and mental health in olderage: a quasi-experimental evaluation of free bus passes in england.J Epidemiol Community Health, 72(5):361–368.

Sobral, T., Galvão, T., and Borges, J. (2019). Visualization of urban mobility data from intelligent transportation systems.Sensors, 19(2):332.

Sobral, T., Galvão, T., and Borges, J. (2021). Knowledge-assisted visualization of multi-level origin-destination flows using ontologies.IEEETransactions on Intelligent Transportation Systems, 22(4):2168–2177.

Suman, H. K. and Bolia, N. B. (2019). Improvement in direct bus services through route planning.Transport Policy, 81:263–274.

Szeto, W., Yang, L., Wong, R., Li, Y., and Wong, S. (2017). Spatio-temporal travel characteristics of the elderly in an ageing society.Travel Behaviour and Society, 9:10–20.

Tang, T., Liu, R., and Choudhury, C. (2020). Incorporating weather conditions and travel history in estimating the alighting bus stops from smart card data. Sustainable Cities and Society, 53:101927.

Trépanier, M., Tranchant, N., and Chapleau, R. (2007). Individual trip destination estimation in a transit smart card automated fare collection system. Journal of Intelligent Transportation Systems, 11(1):1–14.

Wang, R., Das, S., and Mudgal, A. (2020). Patterns of origin destination distributions.

Wong, R., Szeto, W., Yang, L., Li, Y., and Wong, S. (2018). Public transport policy measures for improving elderly mobility. Transport policy, 63:73–79.

Wood, J., Dykes, J., and Slingsby, A. (2010). Visualisation of origins, destinations and flows with od maps. The Cartographic Journal,47(2):117–129.

Yang, H. and Rakha, H. (2019). A novel approach for estimation of dynamic from static origin–destination matrices.Transportation Letters,11(4):219–228.

Yu, C. and He, Z.-C. (2017). Analysing the spatial-temporal characteristics of bus travel demand using the heat map.Journal of TransportGeography, 58:247–255.

Zhao, J., Rahbee, A., and Wilson, N. H. (2007). Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. Computer-Aided Civil and Infrastructure Engineering, 22(5):376–387.