



Revealing personal activities schedules from synthesizing multi-period origin-destination matrices

Haris Ballis, Loukas Dimitriou*

LarB for Transport Engineering, Dept. of Civil and Environmental Engineering, University of Cyprus, 1, Panepistimiou Av., 2109 Aglantzia, Nicosia, Cyprus

ARTICLE INFO

Article history:

Received 27 November 2019
Revised 10 June 2020
Accepted 18 June 2020
Available online 2 July 2020

Keywords:

Activity schedules
Origin-destination matrices
Tours
Trip-chains
Graph-theory
Integer programming

ABSTRACT

Over the last decades, technological advances have allowed the capturing of travel behaviour at large-scale. Despite the unprecedented volume and the variety of personal mobility data, aggregate Origin-Destination (OD) matrices are still the most widespread means to organise and represent travel demand. Nonetheless, standard ODs cannot adequately capture significant elements affecting travel behaviour such as trip-interdependency and trip-chaining, therefore they are not particularly suitable for travel behaviour analysis at person-level. The currently presented modelling framework enables the in-depth study of personal mobility by firstly combining the trips present in OD matrices into home-based trip-chains (i.e. tours) and subsequently into sequences of activities (activity schedules). The above-mentioned process is completed based on advanced graph-theoretical and combinatorial optimisation concepts. The applicability of the methodology is meticulously verified through a large-scale test case where a set of multi-period, purpose dependant ODs is converted into realistic activity schedules able to incorporate more than 99% of the inputted travel demand. The accurate and highly detailed results showcase the significant potential of the proposed methodology to support the comprehensive analysis of travel behaviour at person level.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Motivation and problem statement

Recent technological advances have supported the efficient coordination between transport demand and supply provision and as a consequence have facilitated the emergence of new service models (e.g. Mobility as a Service, Demand Responsive Transport etc.) able to drastically change the mobility motif. Therefore, the re-evaluation of the way mobility has been traditionally modelled and assessed becomes a necessity. Person-centred modelling methodologies such as agent-based and activity-based modelling have been known to the research community for a few decades already but only recently have started gaining significant momentum ([Ben-Akiva et al., 2007](#); [Bradley et al., 2010](#); [Pinjari and Bhat, 2011](#); [Ronald et al., 2015](#); [Cich et al., 2017](#); [Djavadian and Chow, 2017](#); [Bachir et al., 2019](#)). These approaches suggest a shift from the aggregate

* Corresponding author.

E-mail addresses: ballis.theocharis@ucy.ac.cy (H. Ballis), lucdimit@ucy.ac.cy (L. Dimitriou).

modelling paradigm to a disaggregate equivalent where interactions and interrelations between agents can be modelled explicitly. Thus, the need for highly detailed mobility input becomes essential.

Despite the advances in mobility tracking technology and the availability of relevant information in a plethora of data sources (e.g. Call Detail Records, GPS Traces, etc.), the most widely used mean to represent travel demand is still the standard form of Origin-Destination (OD) matrices. ODs represent mobility as the total volume of trips between pairs of locations, often segregated by dimensions such as the purpose of the trips, the transport mode and the time period of departure. Their straightforward structure has facilitated the transferability of results and has established them as the main data exchange format in the transport community. Moreover, their continuous development (Nie et al., 2005; Zhao et al., 2007; Ma et al., 2013; Iqbal et al., 2014; Bonnel et al., 2015; Çolak et al., 2015; Tolouei et al., 2017) indicates that their usability and value will not diminish in the foreseeable future. Even though OD matrices are perfectly suitable for the aggregated representation of travel demand, their format forbids the representation of complicated travel behaviours such as trip-chaining or trip interdependency (Pendyala and Goulias, 2002; McNally and Rindt, 2008). Consequently, they prove inadequate for the in-depth study of complex travel behaviour phenomena (e.g. trip-chains, tours, activity scheduling, etc.). Nonetheless, since typical OD matrices still consist the main input for a variety of transport planning and policy making purposes, approaches aiming at the unveiling of personal travel behavioural patterns within ODs can prove of significant importance.

In the current paper, a novel methodology is proposing the synthesis of individualised activity schedules through the exploitation of the spatiotemporal as well as the trip-purpose information in typical ODs. This is achieved in a stepwise approach. Firstly, an advanced graph-theoretical module (Section 3.3) exploits the spatiotemporal information present in the input ODs in order to enable the identification of all the plausible tours. Subsequently, the available trip-purpose information is used to translate tours into activity schedules. Finally, the combination of activity schedules which best represents travel demand as originally expressed in the input ODs is estimated by an integer optimisation module (Section 3.5).

1.2. Contribution

The deeper understanding of individual travel behaviour requires rich information at a disaggregate level which is not always easily available. This paper proposes a novel methodology for the synthesis of individual activity schedules based primarily on aggregate OD matrices and secondarily on a calibrating distribution of the outputted activity schedules' characteristics. The application of the methodology can prove beneficial to a plethora of cases. For example, nowadays it is technically possible to collect personal travel behaviour data at large quantities (e.g. extensive travel surveys, mobile phone records, GPS traces, etc.) but such data tend to be particularly sensitive (in terms of anonymity requirements) and expensive (Barthelemy and Toint, 2013; Avegliano and Sichman, 2019; E. Ramadan and P. Sisiopiku, 2019). On the other hand, aggregate OD matrices do not usually face these disadvantages but lack in terms of representational detail. The presented methodology can be utilised to enhance the representational value of ODs by converting them into more contextual travel behaviour structures like tours or activity schedules. Another important application of the suggested methodology is the treatment of personal mobility data anonymization. For example, mobile network carriers can provide mobility records of their users/customers but only in the form of aggregated ODs so that individual people's intractability is guaranteed. The proposed methodology provides a valuable alternative for the creation of fully realistic and anonymised travel behaviour information at a disaggregate level. Finally, the currently presented methodological framework can facilitate the transition to disaggregate transport modelling approaches (e.g. agent-based, microsimulation). One of the stated issues hindering their wider adoption is the high-resolution datasets which are often required for the description of agents (Bakker et al., 2014; Srikrishnan and Keller, 2018). As an example, the widely used, agent-based transport simulator MATSim (Axhausen et al., 2016) requires as input the complete travel plans (i.e. location and duration of activities) of the simulated agents. Likewise, traffic microsimulation suites (e.g. SUMO, Vissim, Aimsun Next, etc.) although capable of simulating intricate mobility behaviours such as trip-chaining, they are rarely provided with relevant information. Transport Authorities and planners could also exploit the currently suggested methodology to transform their already available ODs into suitable input for sophisticated agent-based transport simulators enabling the evaluation of complex mobility management and policy making scenarios.

The remaining paper first examines the literature for relevant studies (Section 2), then presents the methodology in detail (Section 3), subsequently proceeds with the validation via a large-scale test case (Section 4) and finally discusses the conclusion and the further steps (Section 5).

2. Literature review

2.1. Identification of patterns in activity scheduling

The study of activity scheduling has been taking place for more than two decades already (Bhat and Koppelman, 1999; Bowman and Ben-Akiva, 2000; Axhausen, 2007). Although, individuals' activity patterns may seem random and unpredictable, recent studies have showcased the ability to actually predict them with great accuracy (Gonzalez et al., 2008; Song et al., 2010; Schneider et al., 2013; Raux et al., 2016). The next section describes the so far suggested approaches for the identification of activity schedules based on different data sources.

2.1.1. Survey data

Prior to the emergence and the wide availability of new and rich information sources, analysts had relied on travel surveys to conduct analysis of activity scheduling patterns. A non-exhaustive collection of indicative examples of such studies is offered here. For instance, [Bowman and Ben-Akiva \(2000\)](#) utilised the Boston 1991 survey to develop an econometric model to impute personal day activity schedules. In 2007, [Lee et al. \(2007\)](#) developed simultaneous, doubly-censored Tobit models to estimate the relationships between household type and structure, time allocation strategies, and trip-chaining patterns, using data from the 2000 Tucson Household Travel Survey. [Nurul Habib \(2011\)](#) developed a random utility maximisation framework for the modelling of dynamic weekend activity scheduling based on information available in the CHASE survey collected for the Toronto area in 2002. More recently, large scale travel surveys, requiring sophisticated analysis approaches, have also become available. For example, [Jiang et al. \(2012\)](#) applied Principal Component Analysis on a large travel survey including more than 30,000 individuals to explore daily activity structures and to cluster them based on socio-demographic information. According to their results, seven to eight groups are adequate for the representative classification of individual activity patterns. Similarly, a network-based approach to identify and categorise activity patterns of individuals was presented by [Zhang and Thill \(2017\)](#). In their research, Zhang and Thill provide a methodology for the clustering of travellers in ‘community structures’ where individuals in the same community tend to interact more intensively compared to agents belonging to different communities. The potential of their methodology to classify large datasets of space-time trajectories was evaluated using 9000 individual travel spanning across Carolina, US.

A widely examined stream of research aiming at the identification of activity schedules is that of Sequence Alignment Methods (SAMs). Although SAMs were originally developed to study DNA sequences, they have been also utilised to study the sequential dependencies between daily activities ([Wilson, 1998; Joh et al., 2002](#)). These approaches attempt to classify activity-chains (usually obtained from travel surveys) into clusters based on their sequencing characteristics and composition. Despite their wide spread, SAMs have been criticised for their inability to capture infrequent activity patterns ([Liu et al., 2015; Saadi et al., 2016](#)). Nonetheless, improvements based on Markovian approaches have been suggested. For instance, [Liu et al. \(2015\)](#) used a profiling method called profile Hidden Markov Models (pHMM) to enable the capturing of the irregular activity patterns. Likewise, [Saadi et al. \(2016\)](#) combined the pHMM method with a population synthesiser to develop a framework capable of assigning activity sequences to all the agents of a population.

Activity scheduling has been also studied under the prism of survival analysis and hazards models ([Ettema et al., 1995; Bhat, 1996](#)). Survival analysis-based methodologies appreciate that the duration between the participation of an individual at the same activity (e.g. work, shopping, leisure, etc.) depends on the elapsed time since the last participation. As an example, [Schoenfelder and Axhausen \(2001\)](#) proposed the use of survival analysis for the identification of rhythmic patterns based on a long-term survey of 316 participants over a course of six years. In the work of [Bhat et al. \(2005\)](#), a sophisticated multivariate hazard model aiming at the examination of the length between successive participations in different activities was developed and applied on a multi-week survey for the cities of Halle and Karlsruhe, Germany. The results indicated distinct weekly rhythms for individuals participating in social, recreational, and personal business activities.

2.1.2. Urban sensing data

The significant role of urban sensing data in the study of travel behaviour has been explored by a large number of researchers ([Caceres et al., 2013; Calabrese et al., 2013; Yue et al., 2014; Çolak et al., 2015; Vlahogianni et al., 2015](#)). Amongst urban sensing data sources, the most widely used for the analysis, clustering and/or estimation of activity schedules are Mobile Phone Data (MPD), GPS traces, and smart-card transit data ([Toole et al., 2015; Anda et al., 2016; Antoniou et al., 2019](#)). Their potential has been evaluated in many studies. For example, [Ebadi et al. \(2017\)](#) constructed spatiotemporal ‘activity-mobility trajectories’ based on a small (37 smart-cards) but detailed smart-card dataset, obtained from students at the University of Buffalo. Their results presented a prediction accuracy between 75 and 88%, showcasing that smart-card data can be utilised for the accurate estimation of activity recognition. A large sample of smart-card data obtained from the London’s public transport network was utilised by [Goulet-Langois et al. \(2016\)](#) for the identification of travel behaviour heterogeneity between public transport users. The researchers firstly inferred a 4-week continuous activity sequence for each of the smart card holders and then clustered them into 11 distinct sequence structures. Sociodemographic information of a small sub-sample allowed them to identify significant connections between the activity sequence structures and the characteristics of individuals. Smart-card data have been also used as input for hidden Markov Chain models. For example, [Han and Sohn \(2016\)](#), relied on smart-card data and land-use information for the transit network of Seoul to impute activity chains using a continuous hidden Markov model. The modelled results yielded plausible and intuitive activity patterns which are also consistent with observed activity patterns.

Between the available urban sensing data, MPD are gradually becoming the main source of travel behaviour information, mainly due to their relative low-cost, large sample size and extended spatial coverage ([Pan et al., 2006; Chen et al., 2016; Ni et al., 2017](#)). Despite their advantages, MPD are not explicitly designed to fuel travel behaviour analyses, therefore they do not include significant travel behaviour dimensions such as the type of activity executed by the mobile phone users. For that reason, analysts have attempted to infer the type of the executed activity mostly through rule-based approaches ([Chen et al., 2014](#)). An attempt to improve the activity type estimation by combining Points Of Interest (POIs) datasets with MPD is presented by [Phithakkittnukoon et al. \(2010\)](#). In Phithakkittnukoon’s research the probability to execute a certain type of activity was calculated based on the number and the type of POIs laying inside each of the modelled areas. A mechanistic approach to synthesise urban mobility profiles through the exploitation of data generated by communication technologies

(i.e. MPD) is presented by [Jiang et al. \(2016\)](#). Jiang et al. utilised MPD to model the location and the duration of primary ('Home' and 'Work') as well as secondary (e.g. 'Other') activities using a rank-based Exploration and Preferential Return (r-EPR) mechanism. Furthermore, the use of MPD derived trip-chains as input for a microsimulation agent/activity-based model has been explored in [Zilske and Nagel \(2015\)](#). The researchers' results indicate that MPD combined with other sources of information such as traffic counts can provide valuable input to simulation models. On a similar stream, [Liu et al. \(2014\)](#) explored and verified the potential of utilising MPD to validate activity-based models. As a last example of MPD-based studies, [Eagle and Pentland \(2009\)](#) relied on Principal Component Analysis to identify the behavioural structure of 100 users, carrying their mobile phones for 9 months. The main aim of their study was to identify a set of characteristic vectors (i.e. patterns), termed as '*eigenbehaviors*', which can approximate the individual's actual behaviour. According to their research, utilising just six '*eigenbehaviors*' can approximate individual's travel behaviour with 90% accuracy.

Unprecedented data availability has enabled the study of travel behaviour phenomena with techniques stemming from the field of Artificial Intelligence. As an example, [Liu et al. \(2013\)](#) employed multiple Machine Learning (ML) algorithms on a dataset covering a year of MPD for 80 users. The supervised ML algorithm was trained with the 2.3% of the locations in the dataset where the respective activities performed at these places were known. According to the researchers, the prediction accuracy of the model reached a remarkable 70% which was further increased to 77% after the application of a post processing algorithm. The use of ML algorithms for the identification of a Markov model's parameters is presented by [Allahviranloo and Recker \(2013\)](#). They employed their methodology and showcased the supremacy of the ML-based methodologies against a standard multinomial logit model. Finally, a data-driven modelling framework for the estimation of human mobility trajectories has been presented by [Pappalardo and Simini \(2018\)](#) where observed MPD data were utilised to construct individual diaries based on an Exploration and Preferential Return methodology. The comparison of their results against observed data showcased the capability of the methodology to accurately reproduce the statistical properties of the observed trajectories. Finally, a prominent methodology providing anonymised and fully detailed activity schedules from MPD is presented by [Lin et al. \(2017\)](#). The authors first utilise an Input-Output Hidden Markov Model (IO-HMM) to infer activity sequences and subsequently apply a Long Short Term Memory (LSTM) deep neural network for the assignment of exact locations to the previously identified activities. The framework presented reasonable performance when 465,000 synthetic activity schedules were assigned in a multi-modal, micro simulator model and the observed traffic and transit counts were compared against the corresponding modelled figures.

The previous section presented a wide range of suggested approaches for the studying of trip-chaining and activity scheduling behaviour. Contrary to the previously presented studies, the novel approach presented here is based solely on aggregated travel demand data (i.e. OD matrices) instead of disaggregate information. The only exception stands for the study of [Ballis and Dimitriou \(2019\)](#) where the authors presented a methodology capable of restructuring multi-period ODs into home-based trip-chains (i.e. tours).

2.2. Trip-chaining in OD matrices

Since the proposed methodology relies on the identification of trip-chaining patterns within ODs for the studying of activity scheduling, it was considered useful to provide some background regarding on the topic. The significance of trip-chaining for travel behaviour analysis has drawn considerable attention over the years ([Thill and Thomas, 1987](#); [Goulias and Kitamura, 1991](#); [McGuckin and Murakami, 1999](#); [Yue et al., 2014](#)). Despite the incapability of OD matrices to represent trip-chaining and trip-interdependency phenomena, many researchers have suggested relevant approaches to incorporate such elements into the OD estimation process. Some of these studies have focused on the exploitation of trip-chaining information obtained from automated data collection sources (e.g. smart-cards) in order to enhance the accuracy of the transit ODs estimation ([Wang et al., 2011](#); [Jun and Dongyuan, 2013](#)). A different stream of methodologies has expressed trip-chains as Markov chains with the purpose to convert data obtained from traffic flows to ODs ([Morimura and Osogami, 2013](#); [Tesselkin and Khabarov, 2017](#)). Additionally, efforts to incorporate trip-chaining information in a dynamic OD estimation framework have been also presented ([Lindveld, 2003](#); [Flötteröd et al., 2011](#)). More recently, [Cantelmo et al. \(2019\)](#), suggested the use of an online dynamic OD estimation framework which combines a departure time choice model with a Kalman Filter to identify correlation between different OD pairs in space and time.

The next section presents the suggested methodology for the exploitation of trip-chaining patterns in multi-period, purpose segmented ODs for the synthesis of disaggregate activity schedules.

3. Methodology

3.1. Overview

The currently presented study, proposes a novel methodological framework for the synthesis of disaggregate activity schedules based on multi-period, purpose-dependant OD matrices. The main principle supporting the proposed framework is the observation that the majority of the population begins and ends their daily activity schedules at home ([Bowman, 1998](#); [Schneider et al., 2020](#)), therefore the required trips to complete these schedules belong into tours.¹ Consequently, most of

¹ It is reminded that this study adopts the tour's definition as a sequence of trips starting and ending at the home location ([Primerano et al., 2008](#)).

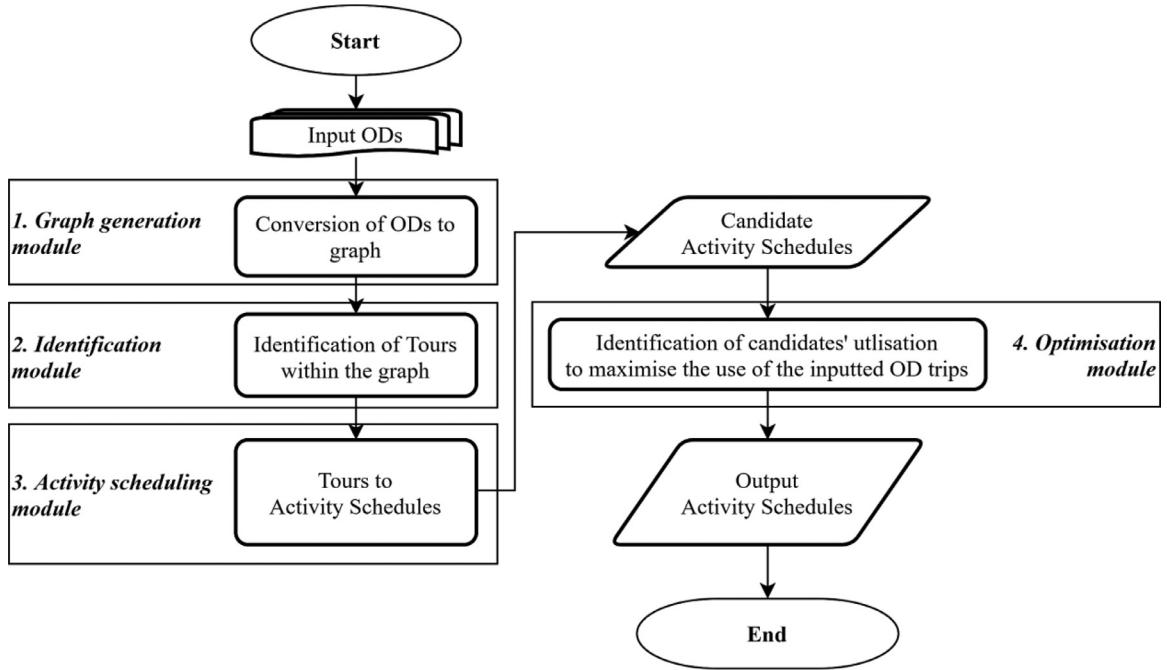


Fig. 1. Flowchart depicting the suggested methodology.

the travel demand captured in ODs should be able to be expressed as a set of disaggregate tours. The aim of the methodology is to identify this set and reconstruct input ODs into the travel demand equivalent tours. Furthermore, the methodology takes this observation a step further by exploiting the, often-included in ODs, trip-purpose information to convert tours into activity schedules.

The identification of activity schedules within ODs is accomplished in a modular fashion. Firstly, the *graph generation module* handles the conversion of the inputted ODs into a suitable graph. Secondly, the graph-theory-based *identification module* completes the identification of all the plausible tours within the graph. Thirdly, the *activity scheduling module* exploits the available trip-purpose information to convert tours into activity schedules. Finally, the linear optimisation-based *optimisation module* calculates the combination of tours whose enclosed trips recreate the inputted travel demand. This modular approach is presented in the following flowchart (Fig. 1).

To sum up, the methodology is accomplished as follows:

- (1) Conversion of input ODs into a suitable graph (Graph generation module – Section 3.2)
- (2) Identification of all the plausible tours in the graph (Identification module – Section 3.3)
- (3) Exploitation of trip purpose information to convert tours into activity schedules
(Activity scheduling module – Section 3.5)
- (4) Identification of the activity schedules' combination which maximises the utilisation of the inputted OD trips (optimisation module – Section 3.5)

For the ease of understanding an example case presenting the expected input OD matrices along with the corresponding output can be found in Table 7 and Table 8 of the Appendix.

3.2. Graph generation module

3.2.1. Conversion of multi-period ODs to a hybrid time varying graph (hTVG)

The first step of the proposed methodology enables the application of advanced graph-theory concepts by converting OD matrices into a suitable graph. A typical representation of a graph can be accomplished by a tuple $G = (V, E)$ where V is the set of vertices (nodes) and E the set of edges (links). This type of representation is very suitable to model situations where relationships between nodes are static. Representing travel demand using graphs is a well-tested approach (Wood et al., 2010). However, most studies either neglect the temporal dimension of demand (Phan et al., 2005) or utilise multiple but isolated networks for the representation of different demand states (Von Landesberger et al., 2016). Nevertheless, travel demand unravels as a highly dynamic phenomenon and therefore it could be more appropriately modelled and analysed as such. In the past few years, intensive research has been allocated on methodologies capable of handling dynamic networks, also known as Time-Varying Graphs (Wang et al., 2019). Following the definition given by Casteigts (2018) a Time-Varying Graph (TVG) can be defined as a tuple $G = (V, E, T, \rho, \zeta)$ where V, E stand respectively for the nodes and the edges of the

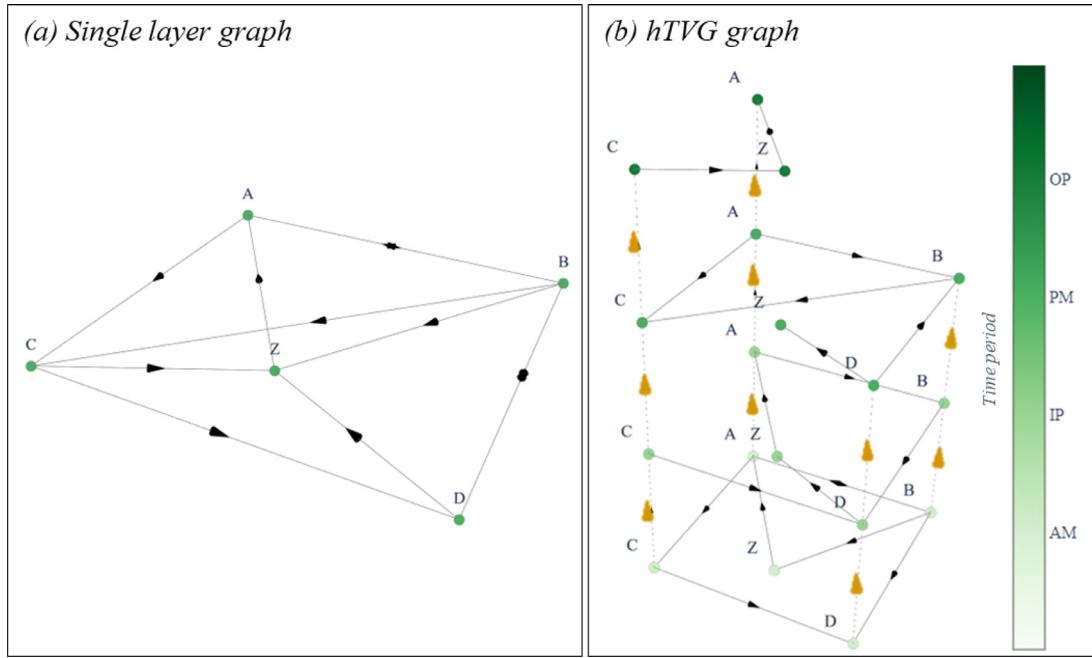


Fig. 2. Conversion of (a) a single layer graph (b) to a hTVG.

network and T represents its lifetime (Time Domain). The dynamic nature of the network is handled by parameter ρ which denotes the presence of each edge $e \in E$ at a given time $t \in T$. Finally, ζ constitutes the latency (i.e. cost) to traverse each edge in E . Based on this, it becomes apparent that dynamic OD matrices can be expressed as TVGs where the zones, trips, time periods and travel-costs constitute the corresponding nodes, links, time domain and latency of the TVG. TVGs have already exhibited their useful properties in numerous studies (Cheng et al., 2003; Ferreira, 2004; Kostakos, 2009) but they can still prove cumbersome to model and manipulate (Casteigts, 2018; Wang et al., 2019). On the other hand, standard static networks have been thoroughly studied for many decades and as a consequence very robust and efficient methodologies have been developed for their analysis. The suggested framework counters the complexity of TVG's by adopting a hybrid solution here referred to as hybrid TVG (hTVG). A hTVG combines the dynamic properties of TVGs with the simplicity of static graphs by expressing the temporal changes as a series of chronologically stacked static graphs, also known as snapshots (Wehmuth et al., 2015). The following section describes the proposed methodology to convert multi-period ODs into a hTVG. In this type of graph, each of the available multi-period ODs is expressed as a separate layer allowing the distinction of trip departures taking place at different time periods. Following this multilayer network format, the spatial characteristics across layers (i.e. the location of nodes on the XY plane) remain stable but the connections between nodes can change, allowing the emergence of variant connectivity patterns across time. In order to simplify the graph, nodes which do not attract or produce any trips are omitted. Nonetheless, without further modification, nodes on different layers are isolated and therefore no paths traversing across different time periods can be formed. To address this issue, nodes representing the same spatial location but in consecutive time periods are connected by *temporal links* (Lin et al., 2016). It must be noted that temporal links do not represent a movement in space neither enclose any cost but are solely used to enable the forward in time transitions. The above-mentioned conversion process is illustrated in Fig. 2. In the single layer graph (Fig. 2a), the spatiotemporal information is expressed on one level. According to this layout, two nodes become connected once a trip takes place between them, regardless of its departure time. Nonetheless, the graph generation module disentangles the temporal information into multiple layers (Fig. 2b) and allows for a more detailed representation of the system. The process is completed by the insertion of the temporal links which can be distinguished by the gold cones notating their direction. Finally, hTVGs can represent more eloquently the temporal variability dimension such as network impedance (travel times, cost, dynamic tolls, etc.) or travel mode availability (e.g. public transport schedules) compared to static graphs. Consequently, any analysis affected by the dynamic nature of the networks (e.g. shortest path identification) is considerably more accurate when executed for hTVGs.

3.2.2. Advantages of hybrid time varying graphs (hTVGs)

The previously presented graph formation exhibits some significant advantages. Firstly, hTVGs achieve the encoding of temporal elements directly into a static-like graph, suitable for the application of efficient graph-theory analysis. The most notable positive effect is that in contrast to single layer graphs, hTVGs do not allow the formation of chronologically inconsistent paths. This is due to the presence of the chronologically directed temporal links which forbid the creation of

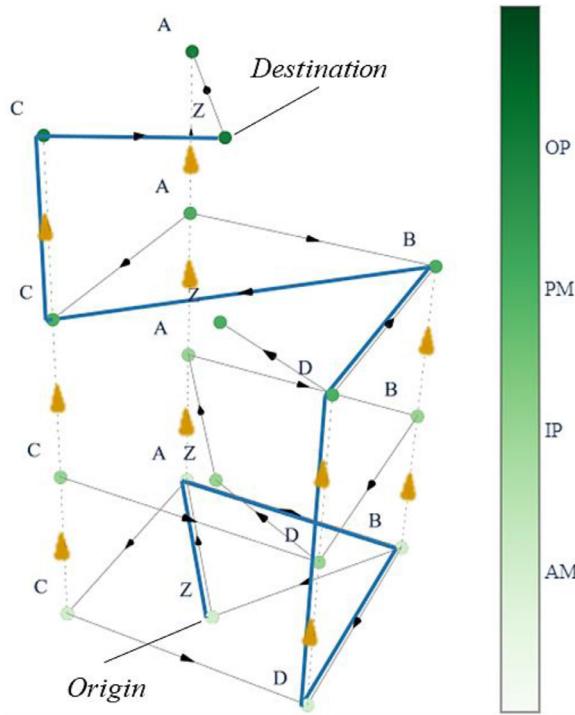


Fig. 3. Formation of a chronologically consistent tour in a hTVG.

unchronological paths. Secondly, the identification of closed paths connecting the same origin and destination (i.e. cycles) is more straightforward in the hTVG case. In graph-theory a cycle represents a closed path with the same origin and destination where no nodes other than the origin can be repeated. Although the cycles identification problem has been thoroughly studied for static graphs, this is not the case for TVGs (Kumar and Calders, 2018). The hTVG format, alleviates this issue by substituting cycles with simple paths connecting the same location in different time-periods. The former two advantages are illustrated in Fig. 3 where a tour from begins from zone Z in the morning (AM) and finishes at night (Off-Peak). As it can be observed, according to the hTVG format, a zone can be connected with itself simply via a path. Additionally, the chronological consistency of the path is guaranteed by the forward in time directionality of temporal links (depicted with gold cones).

The transformation of the single layer network to a hTVG comes with an additional benefit specific to the suggested methodology. As it has been already discussed in one of the authors' previous studies (Ballis and Dimitriou, 2019), one of the key factors affecting the performance of the identification module is the spatiotemporal resolution of the provided network. Based on that study, high-resolution networks result in precise trajectories which can be more easily traced and identified. The quantification of the network resolution and the associated analytical complexity is expressed through the proxy of *network density*. The higher the density of a network, the higher the number of potential paths and consequently the higher the required computation time to identify the full set of candidate tours. Nonetheless, hTVGs present lower density compared to their single-layer equivalents something that leads to reductions of the required processing time. In particular, the network density (d_s) for a directed, single layer network is calculated as the fraction between the actual number of edges E in a graph of V vertices and the number of its plausible edges:

$$d_s = \frac{E}{V(V - 1)} \quad (1)$$

For the equivalent hTVG, the network density (d_h) is calculated as:

$$d_h = \frac{\lambda E + k(V - 1)}{kV(k(V - 1))} \quad (2)$$

where k stands for the number of time periods (layers) of the hTVG and λ ($1 \leq \lambda \leq k$) denotes the increase in the number of links due to the replication of some across multiple layers. Also, the $k(V - 1)$ factor represents the maximum number of temporal links which are required to complete the conversion of a single layer graph to a hTVG. The density of an hTVG is maximised when all nodes are connected with their counterparts in the next layer ($\lambda = k$).

Corollary 1. The density of realistic, single layer transport networks is greater than the density of the equivalent hTVG (i.e. $d_h \leq d_s$) when $k \geq 1 + \frac{V}{E}$.

Proof. Substituting Eq. (1) and Eq. (2) in $d_m \leq d_s$ results to $\frac{kE+k(V-1)}{kV(k(V-1))} \leq \frac{E}{V(V-1)}$. For large networks it can be assumed that $V - 1 \approx V$, therefore $\frac{kE+kV}{k^2V^2} \leq \frac{E}{V^2} \Rightarrow \frac{E+V}{k} \leq \frac{E}{V}$. Finally, solving with respect to k leads to $k \geq 1 + \frac{V}{E}$, which for realistic transport networks holds true since edges are usually at least one order of magnitude more than the nodes (Barabási, 2016). The reduction of density for TVGs has been also experimentally verified by Santoro et al. (2011). \square

3.3. Identification module

3.3.1. Identification of candidate tours

As it has been already discussed, a tour is defined as a sequence of trips originating and ending at the same location. According to the hTVG format, each zone retains its XY position but can appear in multiple time periods (different position on the Z axis). Therefore, a tour can be expressed as a path connecting two nodes representing the same home location but in different time periods. As a result, the identification of tours can be achieved through standard and very efficient algorithms, specifically designed to identify all the paths between a pair of nodes (Sedgewick, 2001). An exception to this procedure is the case of tours starting and ending within the same time period. For these instances, the standard operation of cycles identification is employed. For each zone, the process is applied $\sum_{t=1}^{K(t-1)+K}$ times, where K is the number of time periods that the zone is present in. A visual example is presented in Fig. 4 where zone-Z appears in four time periods. As it can be observed, the identification process has to be executed ten times (six times as a simple path and four times as a cycle identification procedure) to identify all the chronologically ordered tours originating from zone-Z. The full set of plausible tours is obtained by repeating the path identification process for all the zones across all the available time periods.

3.3.2. Search space reduction

Despite the high-performance of algorithms suitable for the identification of paths within graphs, the required time to identify all possible paths between two nodes can grow prohibitively long (Sedgewick, 2001). More precisely, although a single path can be found in $O(V + E)$ time, where V , E stand for the number of vertices and edges respectively, the total number of paths may require significantly more time ($O(V!)$). For the purposes of the current study, the observation that the cost (disutility) of travel has a great impact on the shaping of the daily travel schedule of individuals (Goodwin, 1981; Recker, 2001) can be usefully exploited. For instance, the total travel time of a tour is usually subjected to time or budget constraints, therefore information regarding the users' time-budget can be exploited to discard excessively 'expensive' tours. Moreover, results from travel behaviour analysis have verified that travellers tend to limit the number of trips they execute during a day (Han and Sohn, 2016; Department for Transport, 2017) and only a small percentage of people (less than 2.5%) complete more than five trips a day. The application of sensible travel behaviour-based thresholds can reduce the processing time required to identify all tours within a graph without though discarding frequent travel behaviour patterns. Practically, the reduction of the search space is achieved by the introduction of maximum cost thresholds for various dimensions such as the number of legs in tours, the total travel time, the geodesic distance, the monetary budget, etc. The following figure (Fig. 5) depicts the application of the tour's identification module on the hybrid network which derived from the ODs presented in Table 7 of the Appendix. The identification algorithm is executed with different thresholds regarding the maximum number of legs in the tours. In the first case (a) tours can reach lengths of up to eight legs (excluding the temporal links), while in the second case (b) the threshold is reduced to three legs. The reduction in the search space is significant since the initial 64 tours are reduced to just four. The benefits of imposing such constraints on realistic transport networks can prove even more substantial.

In order to highlight the positive effect of imposing cost thresholds on the required processing time, the computational burden for a hypothetical network of 3000 nodes and 35,000 links is demonstrated in Table 1. As it can be observed, halving the total travel time threshold from four hours to two, reduces the required processing time by at least 96% regardless of the maximum number of allowed legs in tours. The implications are very important because the required processing time to identify all the possible paths between two nodes increases factorially with the maximum length of the path ($O(V!)$). Nonetheless, imposing thresholds can counter this increase and significantly confine the search space.

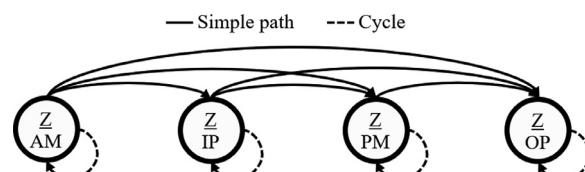


Fig. 4. Presentation of the path identification process. The process takes place for all the progressing time-period combinations.

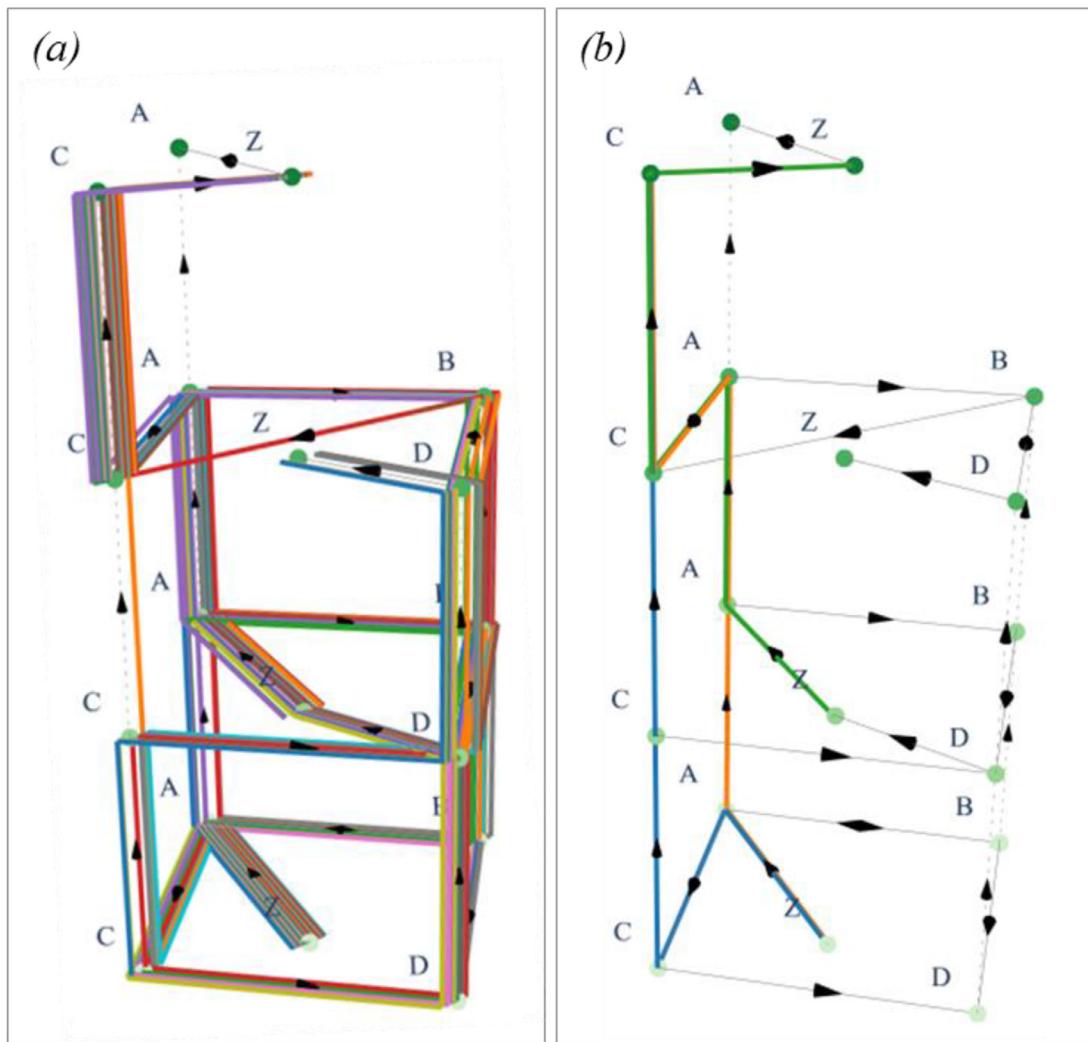


Fig. 5. Identification of all tours originating from zone Z which a maximum number of allowed legs set to (a) eight and (b) three.

Table 1
Required processing time (in seconds) for two total travel time thresholds.

Max tour length (legs)	Processing time (4 h threshold)	Processing time (2 h threshold)	Decrease%
4	34.0s	1.1s	96
5	67.7s	1.2s	98

3.4. Activity scheduling module

3.4.1. Trip-purpose information in OD matrices

The previously presented section describes the required procedure to identify tours within multi-period OD matrices expressed as a hTVG. However, ODs often contain additional information regarding dimensions of travel such as trip-purpose, transport mode, user-group, etc. For the purposes of the presented study, the focus has been placed on cases where trip-purpose information in addition to departure time-period is also provided in the input ODs. These two travel behaviour dimensions are particularly important since they can be utilised to convert tours into detailed activity schedules.

The segmentation of OD trips based on their trip-purpose is a common practice. Trip-purpose is a primary driver of travel decision making which can influence multiple travel behaviour aspects such as destination choice, mode choice, the value of time, etc. The typical categorisation of trips with respect to trip-purpose usually refers to two discrete levels. The first level is with regards to the inclusion (or not) of the traveller's home at either ends of the trip (Home-Based/Non-Home-Based trips). The second level is related to the main purpose each trip is taking place for (work, education, shopping,

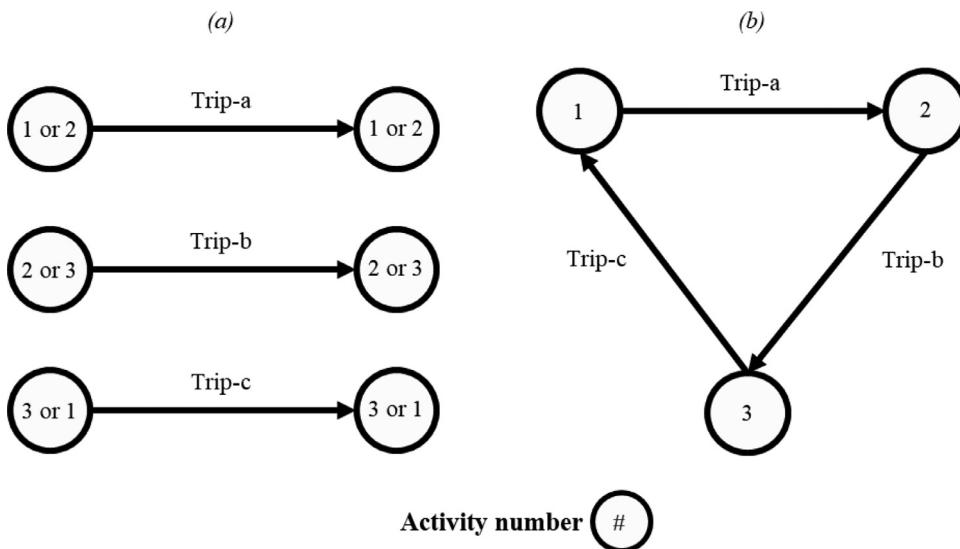


Fig. 6. Chaining of individual OD trips eliminates the ambiguity regarding activity sequencing.
(a) Unchained trips (b) Chained trips.

employer's business, etc.). Nonetheless, a serious limitation for most ODs is that they do not include information regarding the sequencing of activities at the ends of each trip (Ortúzar et al., 2011). For example, a HomeBased-Work (HBW) trip can be used to either express the transition from home to work or vice versa. The methodology presented in the following paragraphs presents an approach to exploit trip-chaining for the inference of the type of activity at the ends of OD trips.

3.4.2. Tours to activity sequences

A basic assumption in travel behaviour theory reads that trips are regarded as the necessary mean to enable the transition from one activity to the next. Therefore, trip-purpose information can be utilised to infer the executed activities at the ends of trips and subsequently enable the conversion of tours into sequences of activities. Tours resulting from the so far presented process contain the required information to enable this conversion. The methodology exploits the fact that activities within a tour take place in a sequential and closed loop fashion, therefore the ambiguity regarding the sequence of the activities can be eliminated (Fig. 6). This is further elaborated through the following example.

Assume a set of four different ODs used to segregate trips according to their trip-purpose, namely:

- Home-Based (HB): The activity at one end of the journey is staying at home ('Home') while at the other end is either:
 - 'Work' (HBW)
 - any 'Other' (HBO)
- Non-Home-Based (NHB): None of the activities at either ends of the trip is 'Home'.
 - If the activity at one end is 'Work' then the trip is classified as (NHBW) while
 - in all 'Other' cases as (NHBO).

Consider also a 4-leg tour which contains two HB and two NHB trips. As described above, traditional ODs do not provide information regarding the sequence of activities at the ends of each trip. Nonetheless, combining the individual trips into tours, allows for the elimination of activity sequencing ambiguity. As noted in the first row of Table 2), a tour cannot consist of a HBW trip followed by a NHBO, then a NHBW and a final returning to home HBO trip because no valid combination of the enclosed activities can be formed. The initial HBW trip can only signify a transition from 'Home' to 'Work' and consequently the next trip should connect 'Work' with the subsequent activity. However, a NHBO trip does not include the 'Work' activity in its definition, hence this trip-purpose sequence is not valid. On the other hand, the rest of the presented

Table 2
Identification of activity sequences from trip-purpose sequences.

Trip-purpose sequence	Activity sequence	Valid
HBW;NHBO;NHBW;HBO	(Home Work), (Other Other), (Other Work), (Other Home)	No
HBW;NHBW;NHBO;HBO	(Home Work), (Work Other), (Other Other), (Other Home)	Yes
HBO;NHBO;NHBW;HBW	(Home Other), (Other Other), (Other Work), (Work Home)	Yes

* The pipe symbol ('|') denotes the 'OR' operator.

Table 3

Definition of an example activity schedule as sequences of various types.

Activity sequence	Locations	Departure time-periods	Departure times
Home; Other; Other; Work; Home	Z; A; B; C; Z	AM; AM; AM; PM	07:45; 08:05; 08:25; 17:30

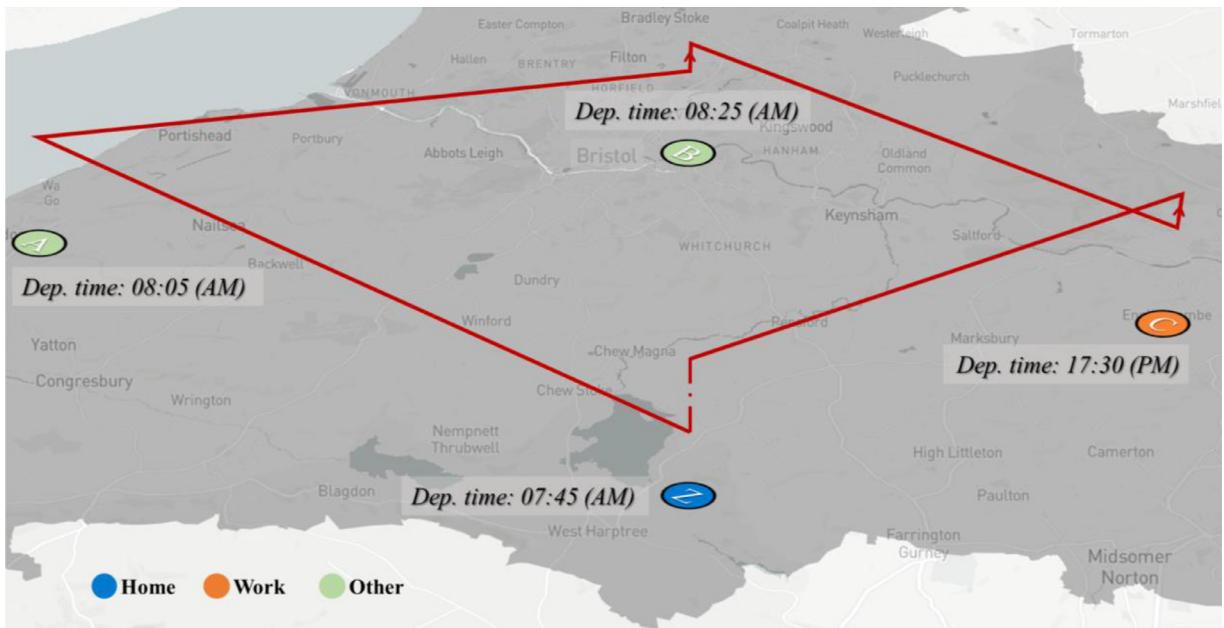


Fig. 7. Visual representation of a typical activity schedule. A person departs from home (zone-Z), completes a series of activities (zones A, B, C) and returns back home later in the day.

trip-purpose sequences are valid and should not be eliminated. The application of this methodology to all the candidate tours (Section 3.3.1), results in a set of candidate activity sequences.

The implications of this observation are significant since it enables the enrichment of typical ODs with information regarding the sequencing of activities for the captured trips. Once the methodology is completed, the individual trips utilised to synthesise tours will be additionally labelled according to the sequence of activities they connect.

3.4.3. Activity sequences to activity schedules

The activity sequences obtained from the so far presented methodology can be further enriched with information regarding the time of departure from each activity. Since the time period of departure for all the inputted trips is known, an estimation regarding the exact time of departure can be attempted. Evidently, the error of this estimation diminishes with the increase of the temporal resolution in the original ODs (i.e. the number of available time periods). The estimation of the exact departure time from each activity can be completed with simple (e.g. uniform distributions) or more refined approaches (e.g. travel survey based), depending on the required level of detail. This estimation can be further improved by considering the travel time between locations. Ultimately, the assignment of the exact time of departure for each trip, allows the conversion of the activity sequences to fully detailed activity schedules. The implications are significant since the initially aggregate OD trips can be now expressed in a much more detailed, decomposed, informative and contextual fashion.

After the assignment of the exact departure time from each activity within the activity schedules, the latter can be fully defined as sequences of (a) visited locations, (b) departure times (or time periods) and (c) activities. An example of a typical activity sequence is depicted in Table 3 where a traveller departs from 'Home' in zone-Z at 07:45 (AM), executes consecutively two short activities of type 'Other' in zone-A until 08:05 (AM) and in zone-B until 08:25 (AM) respectively and finally leaves 'Work' from zone-C at 17:30 (PM) to return back 'Home' in zone-Z. The traveller can be tracked both in space and time since the vertical dimension (z-axis) is used to represent time duration. The example activity schedule is also visualised in Fig. 7.

3.5. Optimisation module

After the identification of the candidate activity schedules, the next step entails the identification of the activity schedules' combination whose consisting trips express the travel demand in the input ODs as closely as possible. Equivalently, the

optimisation objective is expressed as the identification of the activity schedules' combination which minimises the number of non-used trips in the inputted OD trips. This step is referred as optimisation module and is formulated as an integer programming optimisation problem.

3.5.1. Nomenclature

Sets:

C	Candidate activity schedules ($c \in C$)
K	Available time periods ($k \in K$)
P_k	Zone-pairs in each k ($p_k \in P_k \forall k \in K$)
T_{p_k}	The number of trips between each p_k , as recorded in the input ODs
I	Distribution groups ($i \in I$)

Parameters:

$D_c^{p_k}$	Binary variable indicating whether p_k is part of c
G_c^i	Binary variable indicating whether c belongs to i
b_c^i	The probability of c to belong in i
δ_i	Accepted error between the input and the modelled probability for each i

Variables:

N_c	The estimated (modelled) frequency of usage for each c
-------	--

3.5.2. Formulation

Let C be the set of unique candidate activity schedules ($c \in C$). The aim of the optimisation problem is the identification of the frequency of use (N_c) for each c which optimally reproduces the inputted multi-period ODs where K contains the available time periods. In detail, the objective function described in (Eq. (4)) aims to minimize the absolute error between the total number of trips produced by the utilised number of schedules and the trips present in the input ODs by controlling the utilisation of each unique candidate activity schedule $N_c \forall c \in C$. Hard constraint (Eq. (5)) guarantees that the required trips to form the activity schedule s will not exceed the available trips in the input OD matrices. Additionally, constraint (Eq. (6)) assures that N_c does not become negative. The objective function takes its minimum value of zero when the observed and the modelled ODs are identical.

Due to the combinatorial nature of the problem, it is possible that multiple global optima may exist (Redondo et al., 2011) and consequently that more than one combinations of activity schedules can lead to the same optimal objective function value. For this reason, a mechanism to calibrate the optimisation routine towards the identification of a closer to reality solution is required. If a (joint) distribution describing the characteristics of the expected activity schedules (e.g. total travel time, number of activities, modes of transport used, etc.) is available, then this calibrating distribution can be used to shape the output accordingly. To achieve so, each activity schedule within C is assigned the distribution group i ($i \in I$) which belongs to. For instance, in the case where the distribution regarding the count of activities within schedules is known (i.e. share of activity schedules including two activities, three activities, etc.), schedules are assigned the appropriate distribution group i based on the number of the included activities. Such high-level information regarding the characteristics of the expected activity schedules are not particularly difficult to get since they can be available through sources such as traditional travel surveys. The final constraint (Eq. (7)) guarantees that the resulting combination of activity schedules will follow the calibrating distribution. This constraint can be relaxed by the introduction of the term δ_i which allows for a tolerance between the observed and the modelled distribution shares. The optimisation problem is mathematically formulated as:

$$\min Z = \sum_{k \in K} \left(\sum_{p_k \in P_k} \left(\left| \sum_{c \in C} (N_c D_c^{p_k} - T_{p_k}) \right| \right) \right) \quad (4)$$

subject to:

$$\sum_{c \in C} (N_c D_c^{p_k}) - T_{p_k} \leq 0 \quad \forall t \in T, p_k \in P_k \quad (5)$$

$$N_c \geq 0 \quad \forall s \in S \quad (6)$$

$$\left| b_c^i - \frac{N_c G_c^i}{\sum_{c \in C} N_c} \right| \leq \delta_i \quad \forall c \in C, i \in I \quad (7)$$

The above-mentioned optimisation problem is expressed as an integer linear program (ILP) using the optimisation modelling framework Pyomo (Hart et al., 2017) and solved by the CPLEX Branch-and-Bound optimiser (IBM, 2020). The modular nature of the methodology allows the substitution of the currently used Branch-and-Bound algorithm with any other suitable optimisation technique (e.g. Genetic Algorithms, Simulated Annealing, etc.).

Table 4

Definition of available time periods for trips' departures.

Time period	Covered hours	Duration (h)
OP1	00:00 – 07:00	7
AM	07:00 – 10:00	3
IP1	10:00 – 13:00	3
IP2	13:00 – 16:00	3
PM	16:00 – 19:00	3
OP2	19:00 – 22:00	3
OP3	22:00 – 23:59	2

Table 5

Summary of observed ODs.

Trip-purpose	Time period							Total
	OP1	AM	IP1	IP2	PM	OP2	OP3	
HomeBased-Work (HBW)	2313	9840	5335	6632	9660	3739	9,56	3,8475
HomeBased-Other (HBO)	629	2677	15,09	17,86	2645	1061	254	10,561
Non-HomeBased-Work (NHWB)	6	354	643	815	756	121	8	2703
Non-HomeBased-Other (NHBO)	6	147	256	273	277	57	3	1019
Total	29,54	13,018	7743	9506	13,338	4978	1221	52,758

4. Validation test-case

4.1. Experimental design

The assessment of the proposed methodology was conducted over a realistic large-scale case. A set of multi-period input ODs was synthesised by aggregating activity schedules created based on information available in the National Travel Survey (NTS) of UK (Department for Transport, 2017). These activity schedules formed the ground truth for the evaluation, and they are referred as the '*observed activity schedules*'. The decision to synthesise the required ODs by aggregating activity schedules rather than utilising a pre-existing set of ODs was taken in order to allow for the meticulous, one-to-one comparison between the observed and the resulting (*modelled*) activity schedules. In addition, the true potential of the methodology could have been underestimated if based on a set of pre-existing ODs due to inconsistencies of the input rather than inefficiencies of the methodology itself. The locations of the activities taking place within the observed activity schedules were expressed in the 'Lower Layer Super Output Areas' (LSOAs) zoning system. LSOAs are a standard UK census geographic boundary and as of 2011 UK and Wales are divided in 34,753 LSOAs with a minimum population of 1000 and an average of 1500. For the purposes of the current analysis, a zoning system consisting of 470 LSOAs covering the area of Bristol, UK was used. Additionally, trip departures got classified into the seven time periods presented in Table 4. Moreover, for the simplification of the analysis, only activity schedules completed within a single day were included. Nonetheless, this simplification this does not affect the generalization of the method. Finally, the activities described in the survey were grouped in three basic categories, namely 'Home', 'Work' and 'Other'. The first two activity types are self-explanatory while the activity 'Other' is used to express activities not captured by the former two.

The distribution of the produced activity schedules in terms of departure time periods and total travel time is presented in Fig. 8. The profile of departure time periods refers to the sequence of time periods under which trips within activity schedules take place. At this point it should be noted that the exact ending time of each activity was assigned by a uniform distribution based on the time-period of departure for each associated trip. Moreover, trip durations were calculated based on the distance matrix provided by the OSRM project (Ozimek and Miles, 2011). This distribution is important since it was also used for calibration purposes in order to shape the output towards the observed activity schedules.

Ultimately, the complete observed activity schedules set consists of 24,649 unique schedules which once aggregated result in 52,758 total OD trips, roughly corresponding to 10% of the car trips taking place on an average day in Bristol (UK National Trip End Model). The translation of activity schedules to OD trips is possible because the location and the starting time of each activity is known, therefore the required trips to connect these activities can be inferred. These trips are summarised in a set of ODs referred as the '*observed ODs*'. Trips in the observed ODs are segmented in seven time periods and four trip-purpose groups. Their summary is presented in Table 5 while their visual representation as a hTVG network is depicted in Fig. 9. As it can be noticed, the number of edges between the different layers of the network varies considerably. This is expected since travel demand is not usually uniformly distributed across the day. On the other hand, the number of visited nodes remains generally stable, indicating that the number of at least once visited locations does not vary significantly during the day.

Based on the discussion made in Section 3.3.2, the identification of all the possible tours and activity sequences can prove a remarkably time consuming process. Nonetheless, the available search space can be considerably reduced by the introduction of cost thresholds. For the purposes of the validation, two thresholds were implemented. The first threshold

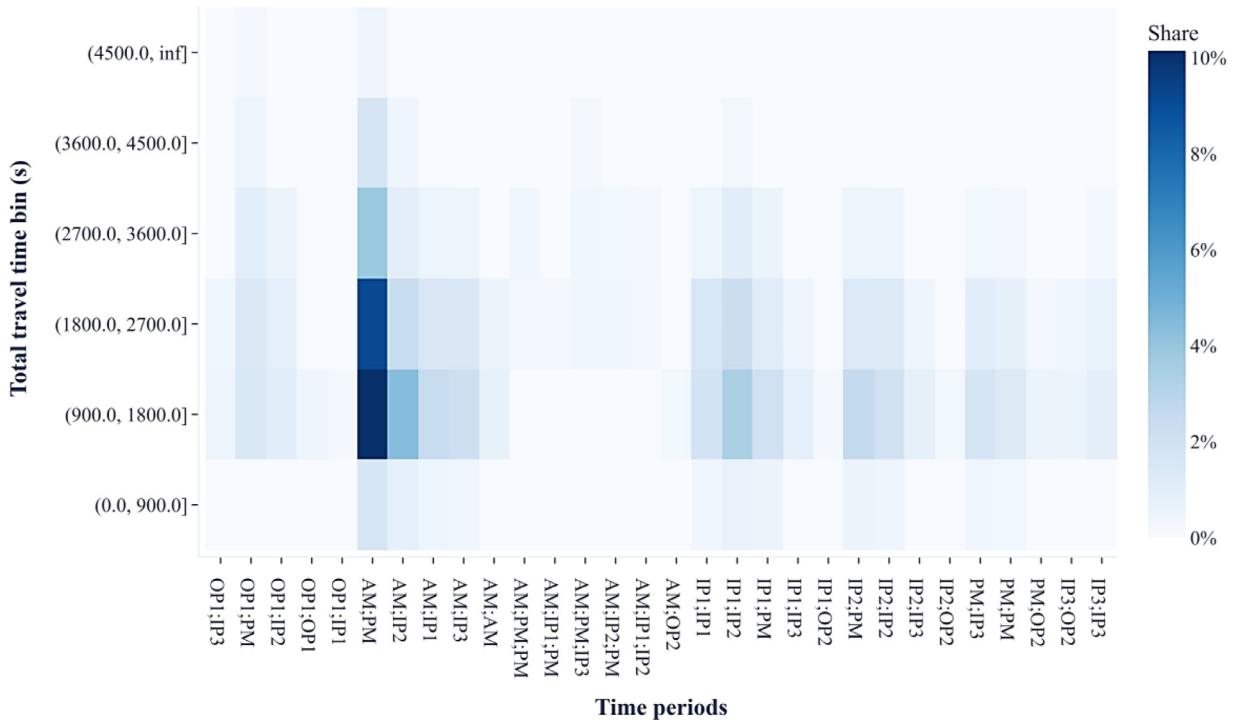


Fig. 8. Joint distribution between the departure profile and the total travel time of the observed activity schedules.

excluded activity schedules with more than five activities while the second excluded activity schedules whose total travel time exceeded two hours. The total share of excluded schedules is 0.72% and 0.23% for the first and the second thresholds respectively. As it can be noted, the total share of the discarded schedules is arguably insignificant, therefore, their exclusion cannot have a substantial effect on the accuracy of the output. On the other hand, the reduction of the search space due to the enforcement of these thresholds is significant and allows for the execution of the process in reasonable time. As an indication, an attempt to complete the identification module without imposing cost thresholds was not successful even after 48 h of processing time. In contrast, the enforcement of thresholds allows for the completion in three hours.

4.2. Discussion of the results

The application of the suggested methodology on the multi-period and purpose-dependant observed ODs resulted in 24,518 modelled activity schedules. As it will be showcased in the next section, the output represents very realistically the input. In terms of performance, the whole process was executed in approximately eight hours on an eight-core i7 processor with 16 GB of available RAM. The optimisation module accrued for 38% (around three hours) of the total processing time while the rest 64% was devoted to the optimisation module. It should not be disregarded that problems of combinatorial nature like the one presented above can often prove particularly cumbersome even with state-of-the art methodologies and high-end computing resources (Klotz and Newman, 2013). Therefore, the previously mentioned solving times can be considered very satisfactory. Since the process is highly parallelisable, additional processing time reduction can occur if computational systems with many threads are to be utilised.

The following section aims to validate the methodology both at an aggregate as well as at a disaggregate level. The aggregate level assesses the ability of the methodology to identify a combination of activity schedules able to represent travel demand as captured in the inputted ODs. The disaggregate level of validation focuses on the assessment of the outputs' realism and entails the one-to-one comparison between the observed activity schedules and the modelled ones.

4.3. Aggregate-level validation

4.3.1. Comparison of ODs

The first level of the methodology's assessment includes an aggregate-level comparison between the observed (input) ODs against the modelled. The 24,518 identified modelled activity schedules utilise more than 99.2% of the total trips present in the observed ODs. To validate this, the comparison between the observed and the modelled ODs by time period of departure and trip-purpose is presented in Table 6. As it can be noticed, the differences between the compared ODs are minimal.

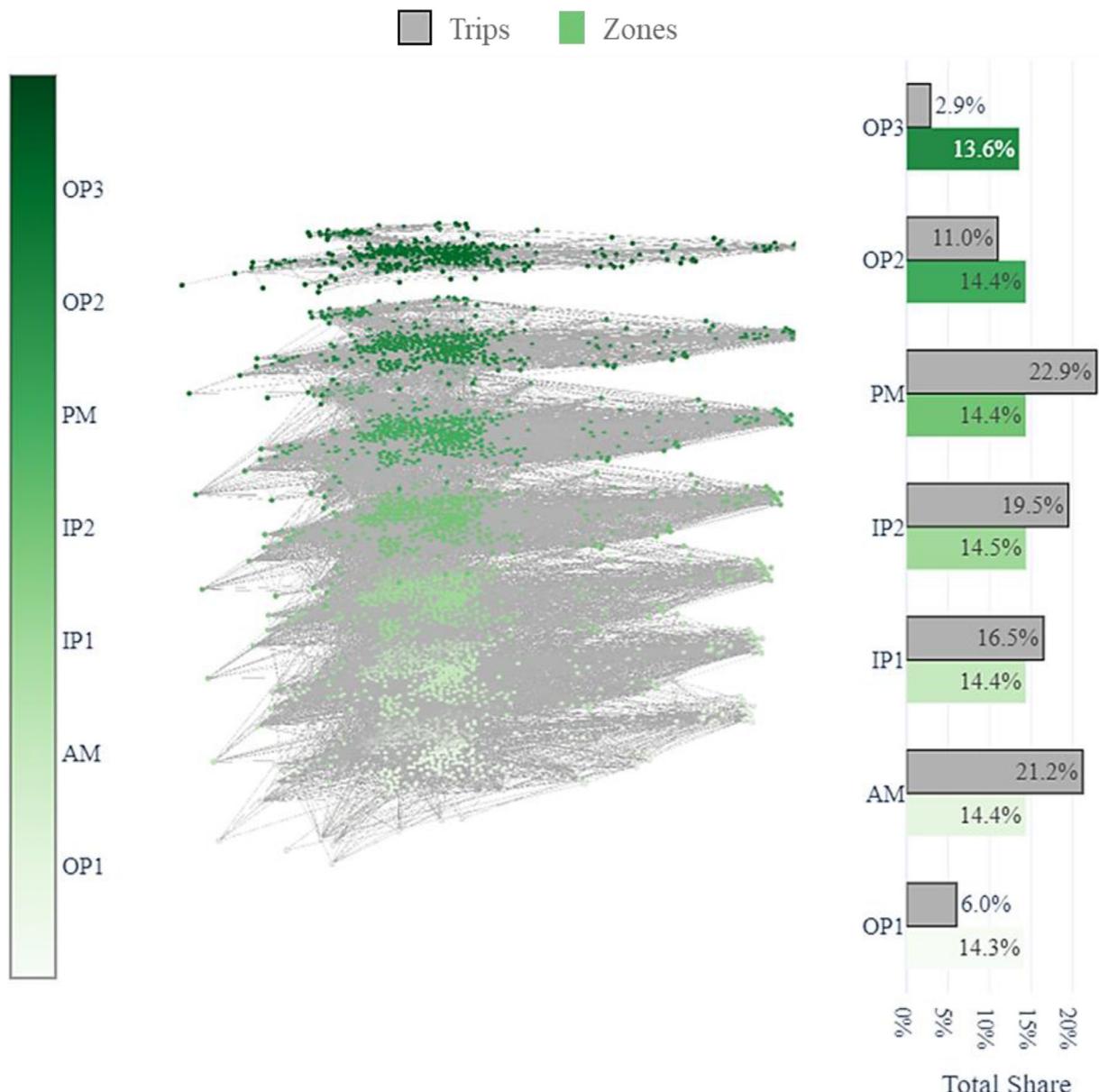


Fig. 9. The multi-level visual representation of the multi-period observed OD matrices as well as the distribution of person trips and visited zones between time periods (layers).

The accuracy of the methodology is also examined in the scatter diagram of Fig. 10. The size of each point represents the number of trips between pairs of locations. As it can be noticed, the number of missing trips (orange points) is significantly lower compared to the number of the observed trips. This can be further verified by the minor error terms visualised in the accompanying histograms presenting (in a logarithmic scale) the total origins and destinations from and to zones.

Finally, the conversion of a multi-period ODs to individual activity schedules is visualised in the two following figures (Fig. 11 and Fig. 12). In particular, Fig. 11 depicts the observed ODs where colour shading and vertical positioning are used to differentiate trips departing at different time periods. Darker tones and higher elevated trips indicate departures later in the day. On the contrary, Fig. 12 presents the completion of the identified modelled activity schedules with each schedule being represented by a different colour. As it can be observed, the majority of trips has been utilised to form activity schedules.

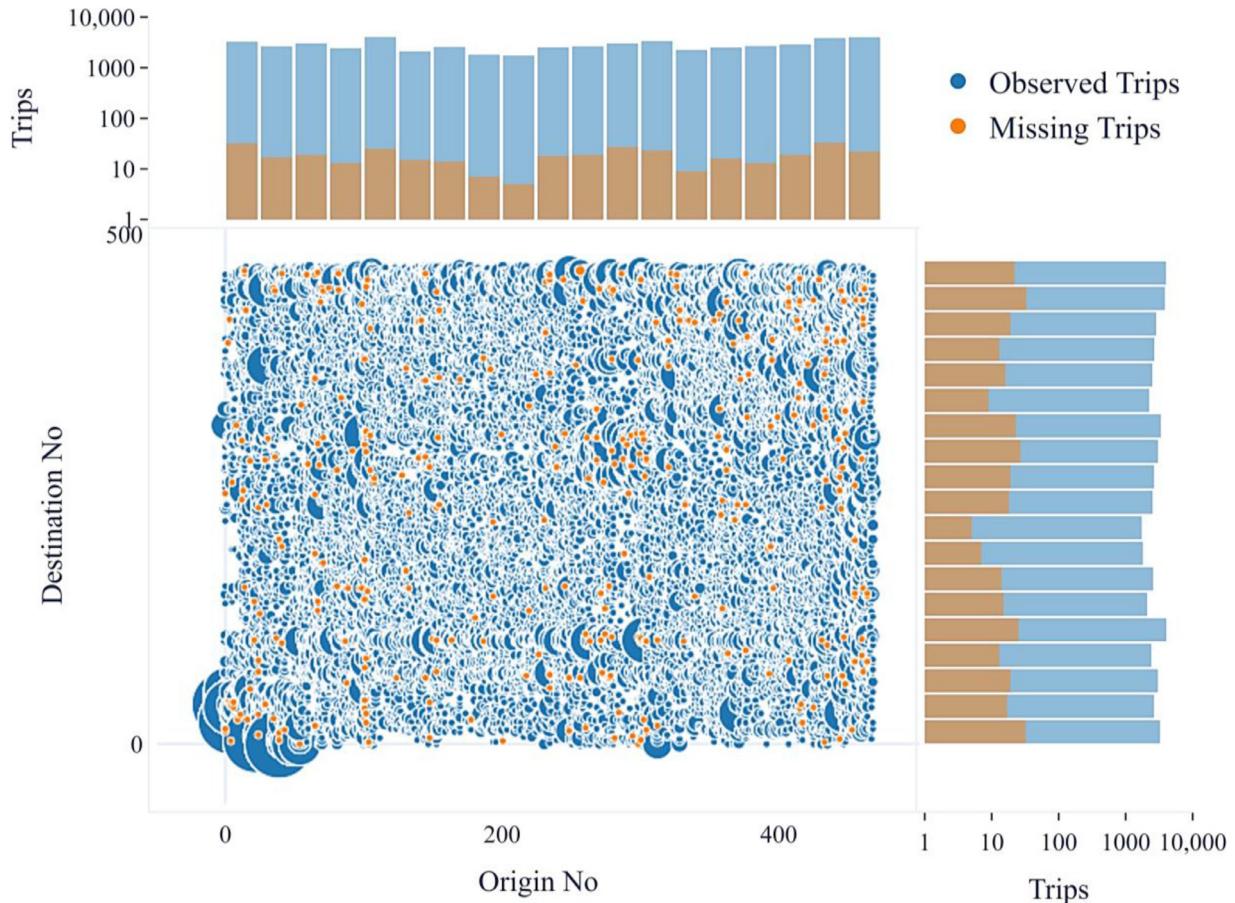
4.3.2. Comparison of high-level distributions

As it has been already discussed in Section 3.5, the activity schedules' combinations which can optimally recreate the inputted travel demand may be numerous. In order to guide the optimiser towards a solution closer to reality, the distribution of activity schedules' high-level characteristics (presented in Section 4.1 and Fig. 8) was enforced as a constraint.

Table 6

Absolute and percentage difference between the observed and modelled ODs.

Purpose	OP1	AM	IP1	IP2	IP3	PM	OP2	Total
HBW	16 (0.7%)	55 (0.6%)	32 (0.6%)	33 (0.5%)	43 (0.4%)	21 (0.6%)	7 (0.7%)	207 (0.5%)
HBO	2 (0.3%)	13 (0.5%)	12 (0.8%)	12 (0.7%)	11 (0.4%)	3 (0.3%)	2 (0.8%)	55 (0.5%)
NHWB	0 (0.0%)	5 (1.4%)	14 (2.1%)	15 (1.8%)	18 (2.3%)	1 (0.8%)	0 (0.0%)	53 (1.9%)
NHBO	1 (14.3%)	2 (1.3%)	9 (3.4%)	9 (3.2%)	7 (2.5%)	2 (3.4%)	1 (25.0%)	31 (3.0%)
Total	19 (0.6%)	75 (0.6%)	67 (0.9%)	69 (0.7%)	79 (0.6%)	27 (0.5%)	10 (0.8%)	346 (0.7%)

**Fig. 10.** Comparison between the number of person trips in the observed and the modelled ODs.

The bar chart presented in Fig. 13 validates the accurate enforcement of the above-mentioned constraint since only minor, expected discrepancies between the share of the observed and the modelled distribution groups can be noted. Moreover, since the absolute number of the observed activity schedules (24,649) is almost equal to the corresponding number of the modelled schedules (24,518), it can be deducted that the absolute number of activity schedules within each of the compared distribution groups is very similar.

4.4. Disaggregate-level validation

The results presented in Section 4.3, validate the capability of the presented methodology to produce activity schedules able to retain the aggregate-level characteristics of the observed activity schedules and at the same time represent the total travel demand as described in the observed ODs. Nonetheless, the complete validation requires the comparison of the schedules at a disaggregate-level.

4.4.1. Comparative dimensions

As a first comment, the distance between the number of modelled activity schedules (24,518) and the observed ones (24,649) is particularly low (0.5%). Despite this low difference, the output was also validated against the individual char-

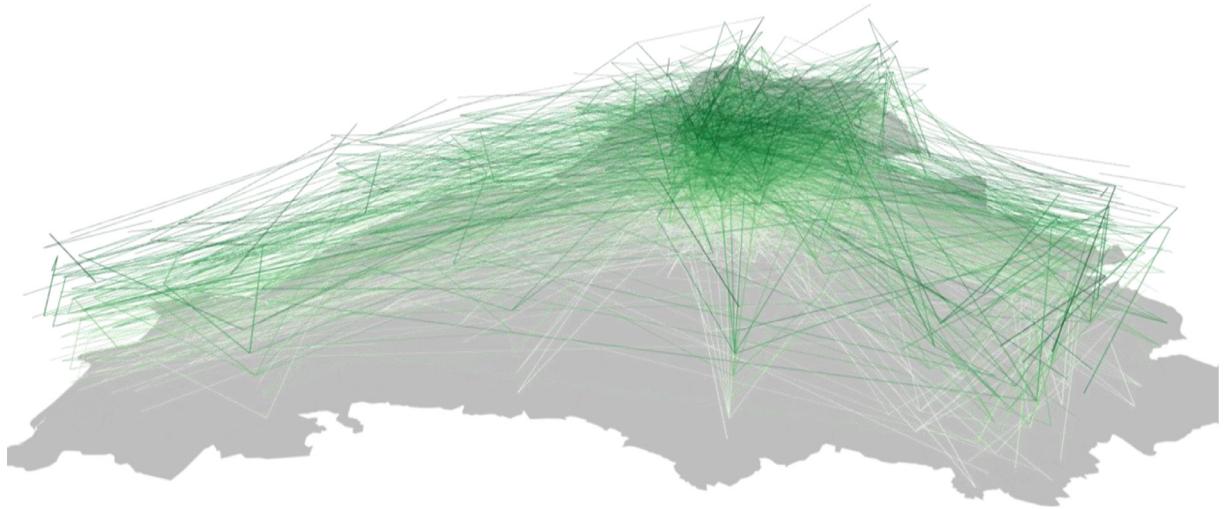


Fig. 11. The multi-period ODs; darker colours indicate person trips departing later in the day.

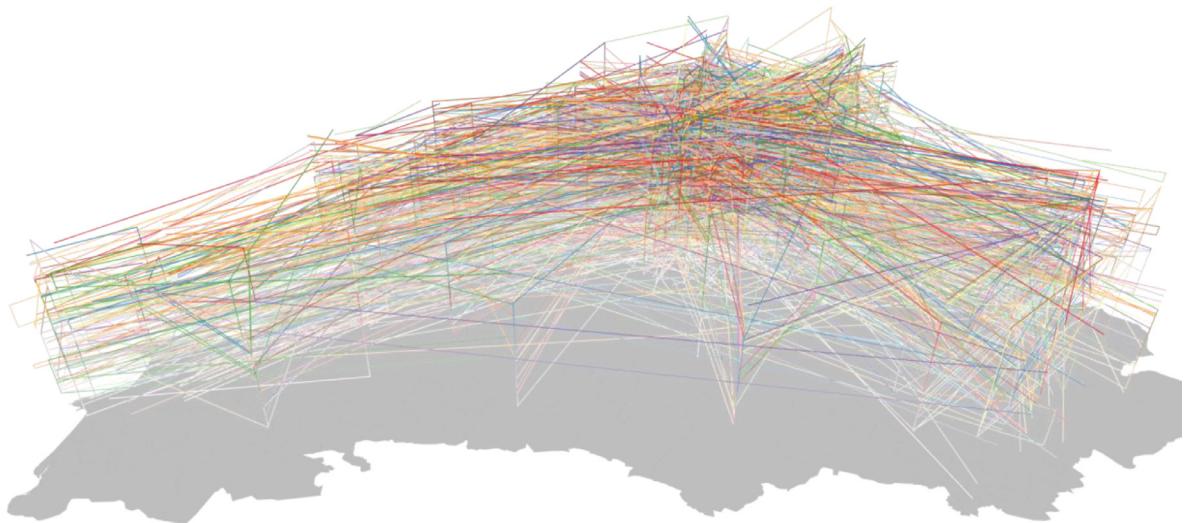


Fig. 12. Individual colour-coded activity schedules.

acteristics (*comparative dimensions*) of each schedule (Section 0). The comparative dimensions which were evaluated correspond to the sequences of (a) the visited locations, (b) the departure time periods and (c) the activity types of each schedule.

For reasons of visual clarity, the onwards analysis focuses on the four most common (out of 24) activity type sequences and the ten most common (out of 256) departure time sequences while the rest of the sequences are classified as 'Rest'. More information regarding this classification can be found in [Table 9](#) and [Table 10](#) of the Appendix. Additionally, the relevant 'Home', 'Work' and 'Other' activity types have been shortened to their initial ('H', 'W', 'O'). Finally, due to the high number of observed location sequences (e.g. Zone-Z; Zone-A; Zone-Z), the corresponding results are grouped in bins of 2000 location sequences each.

4.4.2. One-to-one comparison

To further support the validation process, the percentage of the unmatched observed activity schedules is presented in figure [Fig. 14](#). An observed activity schedule is considered as '*unmatched*' when it cannot be paired with an equivalent modelled schedule. The matching process takes place without replacement, therefore the same modelled schedule cannot be assigned to more than one observed schedules. Depending on the number of the simultaneously considered comparative dimensions, the percentage of the unmatched schedules varies significantly. As it can be noted, the main dimension contributing to the misalignment between the observed and the modelled schedules is the combination of the location and

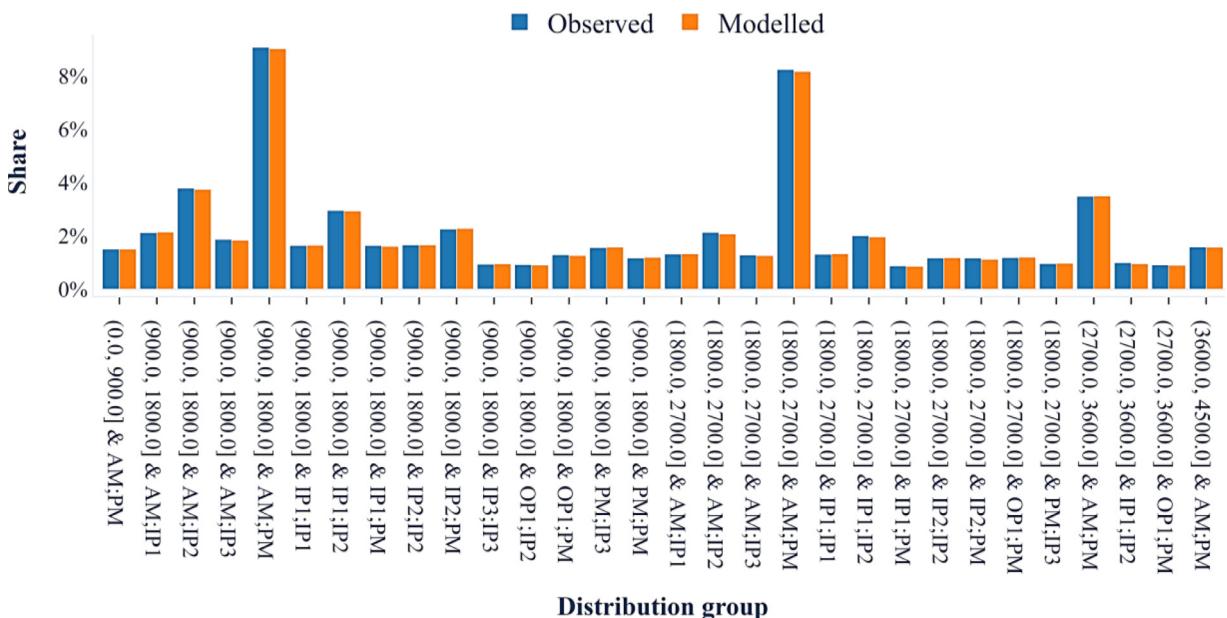


Fig. 13. Comparison of the 30 distribution groups with the largest share between the observed and the modelled activity schedules.

the time period sequences. When examined in isolation, these two sequences attribute for 2.1–2.4% of the discrepancy but their simultaneous examination results in a misalignment of 9.51%. Finally, the percentage of unmatched observed schedules when the comparative dimensions are altogether considered does not exceed 9.55%. This encouraging result provides strong evidence regarding the capability of the methodology to reproduce realistic multidimensional activity schedules based on limited aggregate input.

The next section delves into the distribution of error (i.e. mismatch) amongst the comparative dimensions. Firstly, the distribution of error for each comparative dimension separately is depicted in Figs. 15 to 17. Blue bars represent the share of observed distribution groups while orange bars the percentage of unmatched schedules. As it can be noticed, the error term is distributed almost proportionally between the different groups for all the comparative dimensions. The low percentage error presented in Fig. 15 can be attributed to calibration distribution which controlled the number of schedules within each time-period sequence group. Up to some extent, this also holds true for the low error term presented in distribution

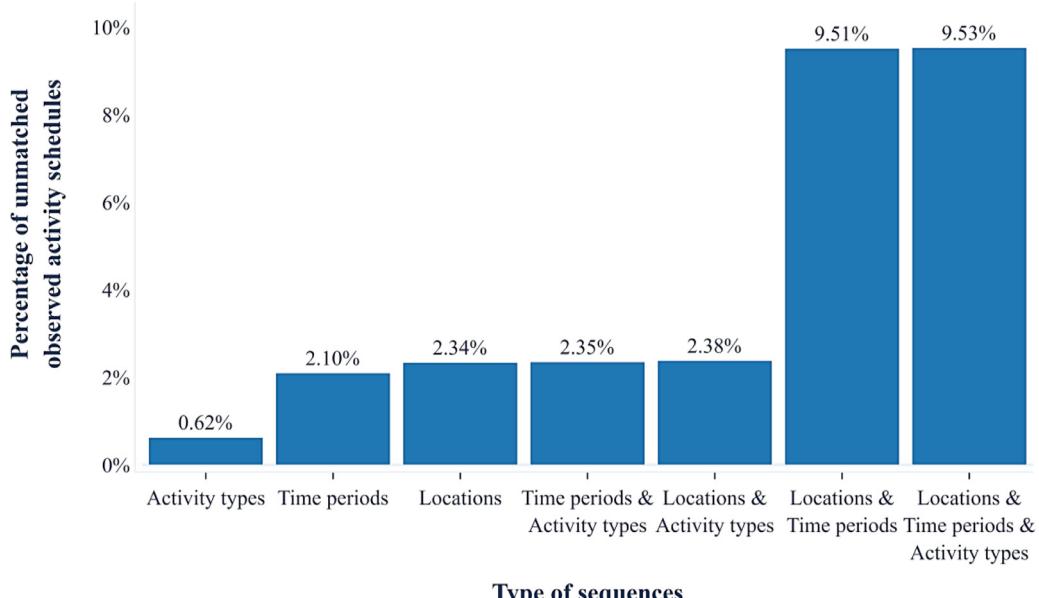


Fig. 14. Examination of the comparative dimensions on the accuracy of the.

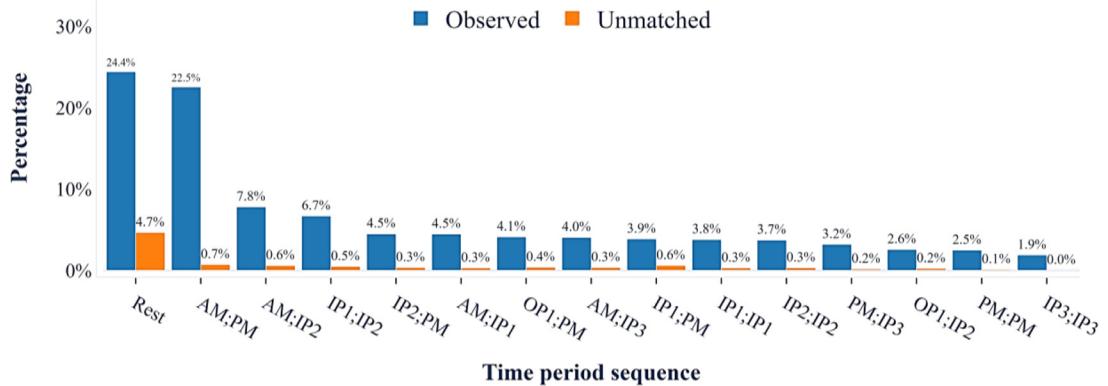


Fig. 15. Percentage of unmatched activity schedules for the departure time periods sequence comparative dimension.

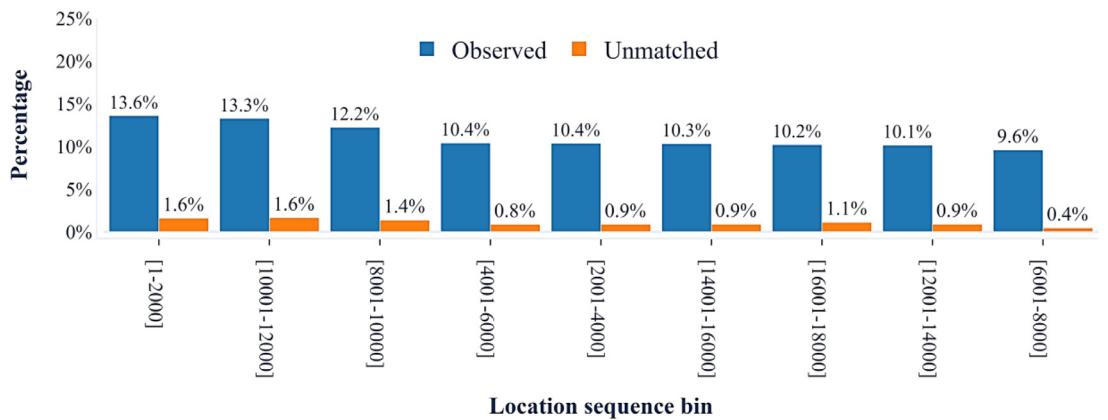


Fig. 16. Percentage of unmatched activity schedules for the location sequence comparative dimension.

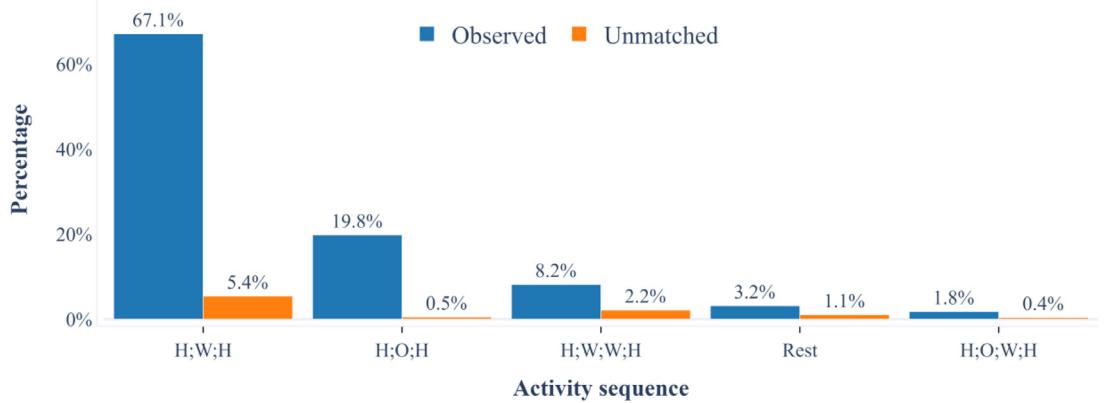


Fig. 17. Percentage of unmatched activity schedules for the activity type sequence comparative dimension.

of error within the location sequence groups (Fig. 16). Although, the exact number of schedules following each location sequence is not known, however, the provided information regarding the total travel time between locations improves the quality of the output. Interestingly though, the methodology has accomplished a particularly accurate solution even for the unconstrained dimension of activity sequences (Fig. 17). This is particularly important because the location and the activity type sequences have been endogenously estimated without relying on any calibration process. Focusing on Fig. 17 reveals that the error term is more notable for complex activity schedules (i.e. schedules including more than three activities) which are nonetheless infrequent and do not significantly affect the accuracy of the output. In cases where the activity

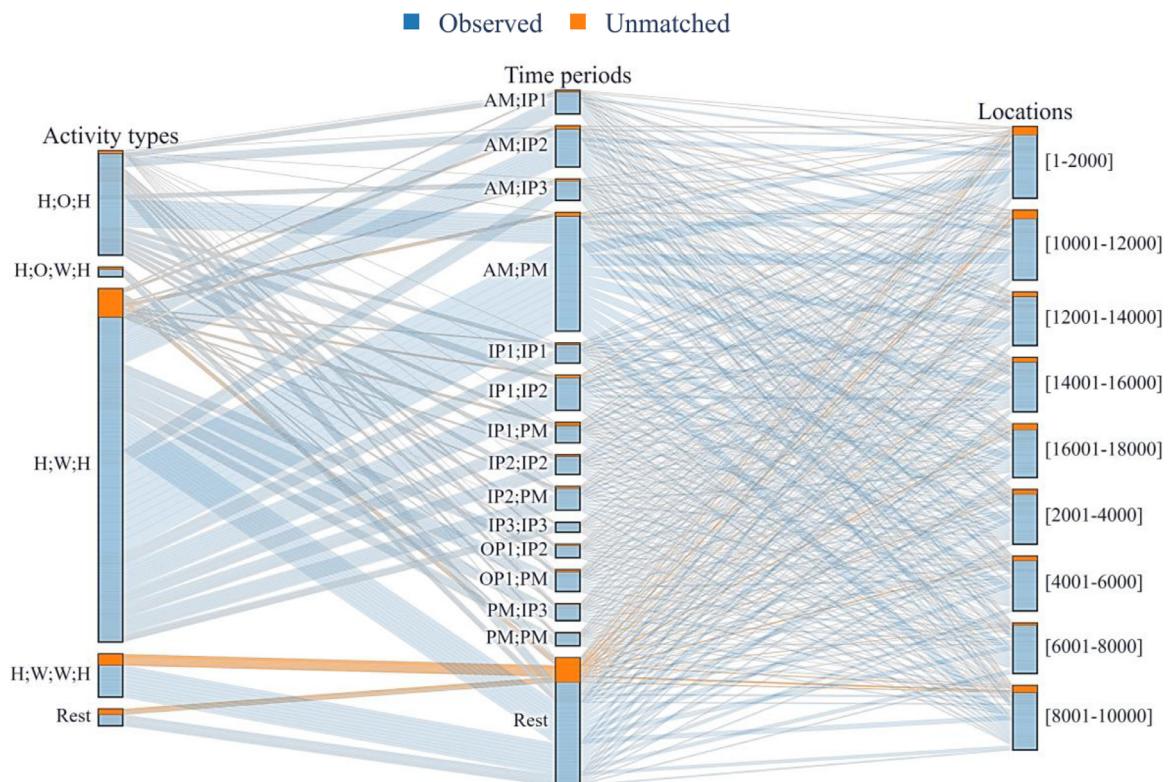


Fig. 18. Presentation of the unmatched activity schedules between the observed ones.



Fig. 19. Distribution of participation in different activities throughout the day.

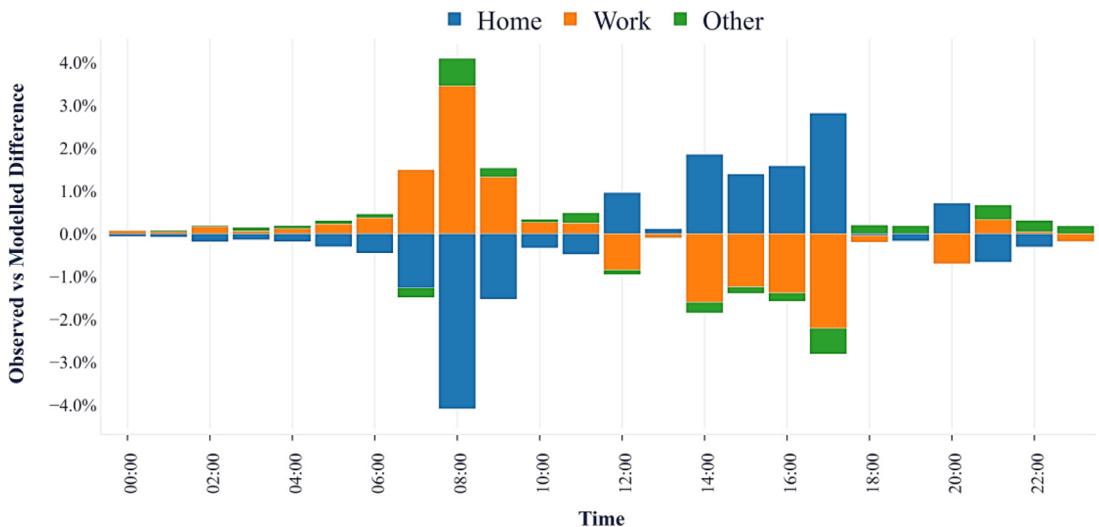


Fig. 20. Percentage difference between the observed and the modelled participation for different activities throughout the day.

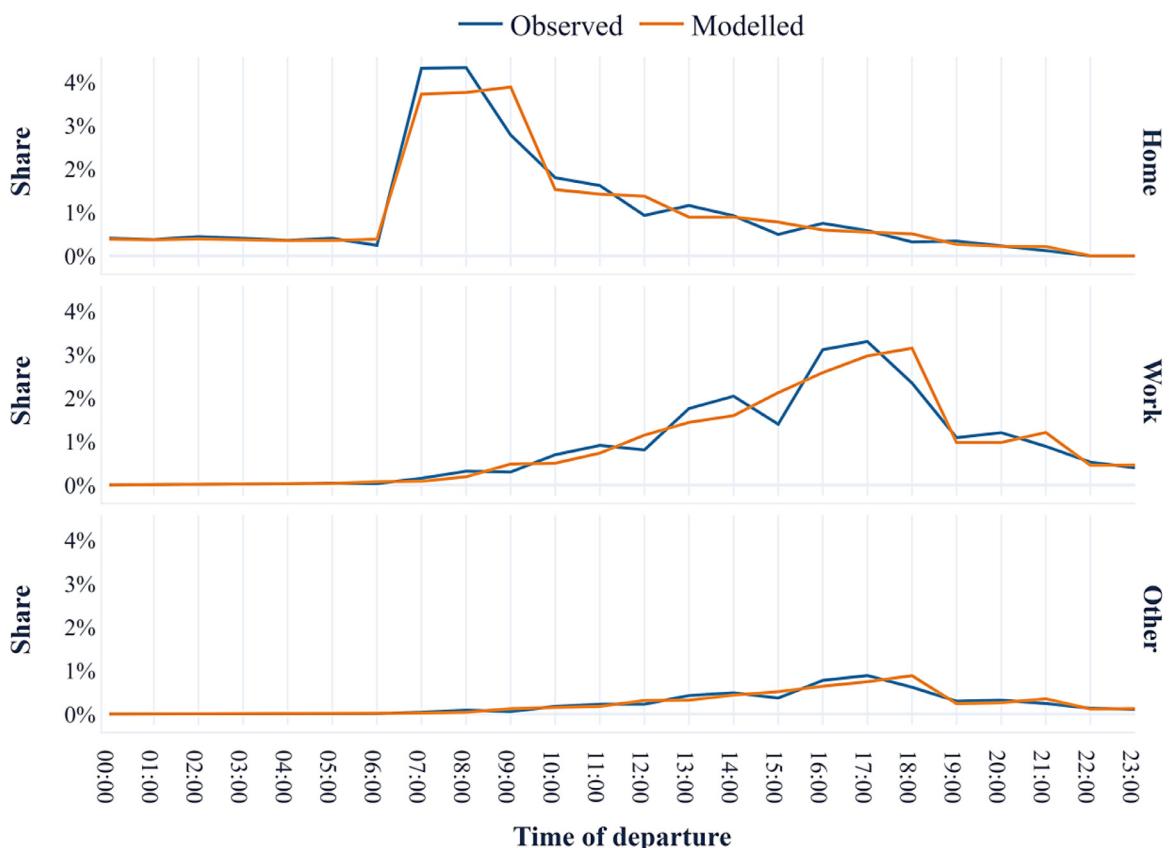


Fig. 21. Departure profiles for the available activity types.

type sequencing is an important factor of the analysis, the researcher should attempt to incorporate such information in the calibration distribution.

Finally, the presentation of the unmatched activity schedules is visualised through the parallel categories diagram of Fig. 18. In this diagram each individual activity schedule is presented as a string crossing through its defining dimensions. The activity schedules which were not perfectly matched are clearly depicted. This visual representation emphatically showcases the high accuracy of the suggested methodology since only a relatively small percentage (9.53%) of the output does not fully comply with the considerably complex input.

4.4.3. Activity participation profiles comparison

The combination of individual trips within ODs into tours and subsequently in activity schedules provide information for further descriptive behavioural analysis such as the participation of the population in different activities. Fig. 19, depicts the distribution of activities taking place in the studied area over the course of a day while Fig. 20 presents the percentage differences between the observed and the modelled activity schedules in terms of their activity participation profiles. As it can be noticed, the comparison between the corresponding distributions assures that the methodology can replicate the observed patterns with great fidelity. This statement is based on the low relative errors (less than 4%) between the observed and the modelled figures.

4.4.4. Departure time profiles comparison

Another set of comparisons between the observed and the modelled activity schedules revolved around their departure time profiles. As it can be noticed in Fig. 21, the proposed estimation framework has managed to replicate the trend of departures between the observed and the modelled schedules without significant discrepancies. Peaks and troughs arise almost at the same time, while the rates of departures are generally similar. Considering that the assignment of the exact departure time for each trip was based on a uniform distribution, it becomes apparent that the methodology can accurately estimate realistic activity schedules even with limited input. In the case where the duration of activities plays a crucial factor to the analysis, the researcher can incorporate relevant information as optimisation constraints.

4.4.5. Duration of activities comparison

The final stream of comparative analysis emphasised on the duration of activities within the observed and the modelled schedules. Fig. 22 depicts the distribution of the duration for the available activity types classified into bins of 1-hour duration. As it can be noticed, the percentage difference for most of the cases is below 0.5% with only a couple of exceptions

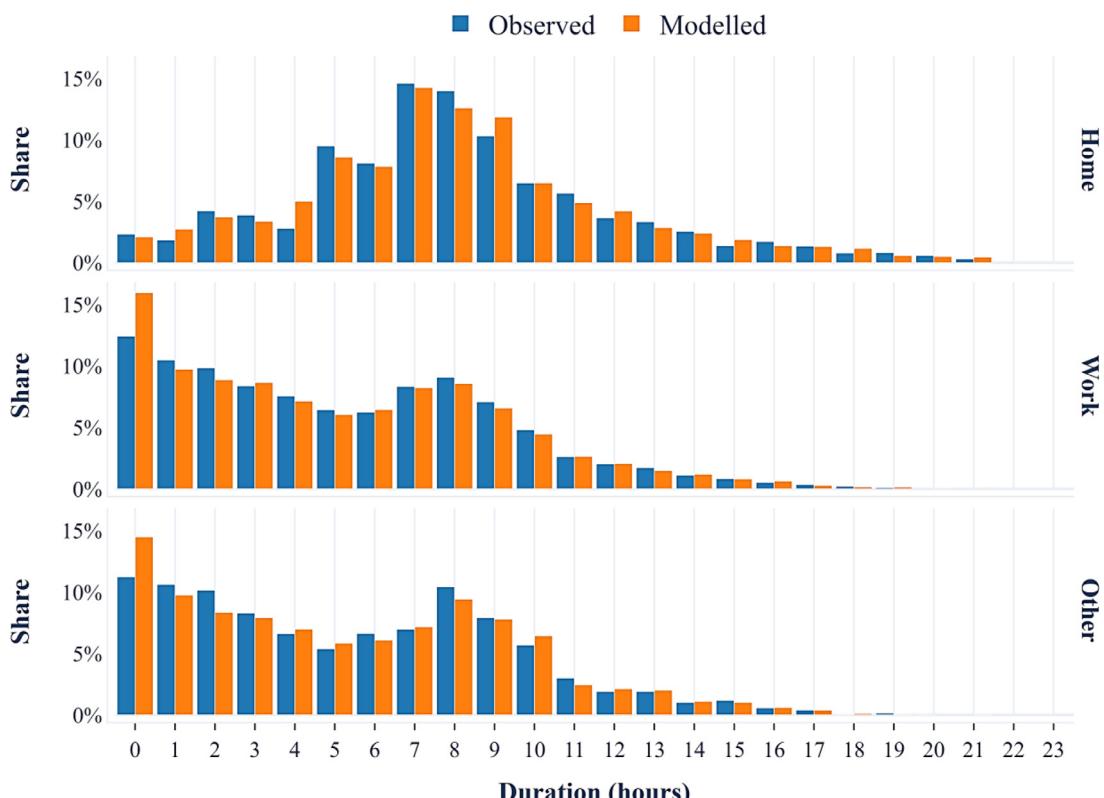


Fig. 22. Duration profiles for the available activity types.

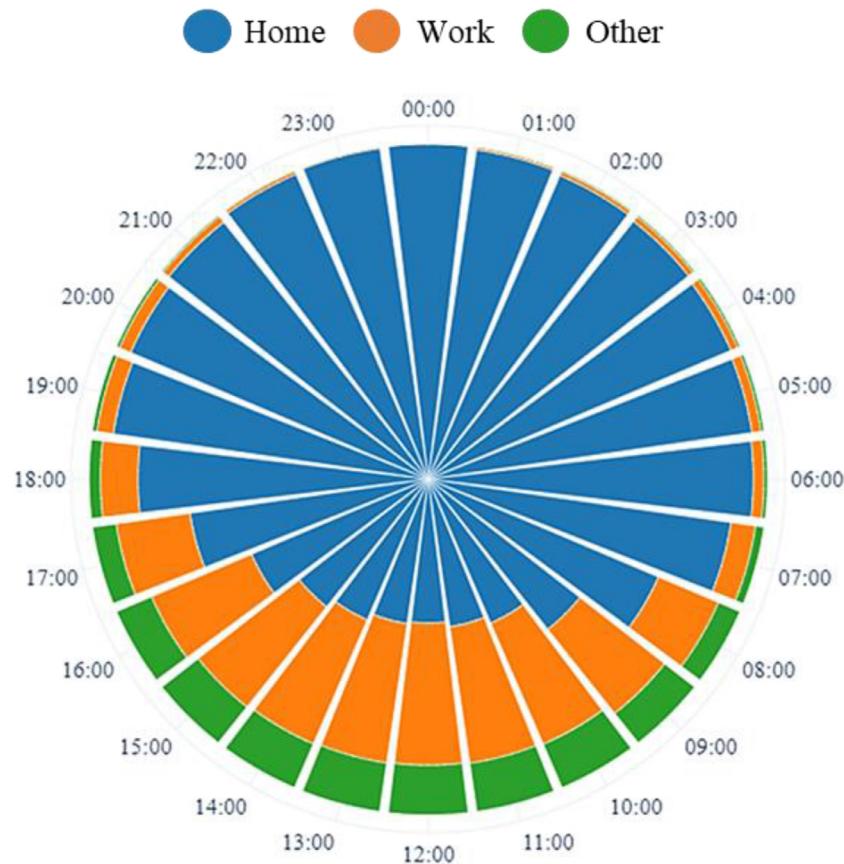


Fig. 23. Profile of activity participation for the studied urban area.

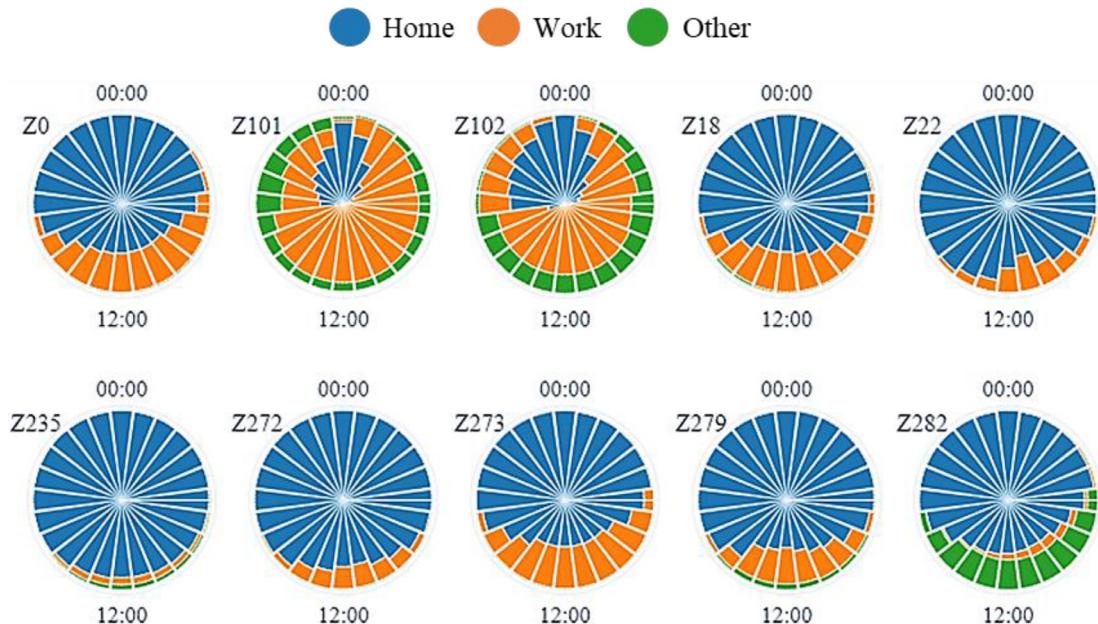


Fig. 24. Profile of activity participation for a set of sampled zones.

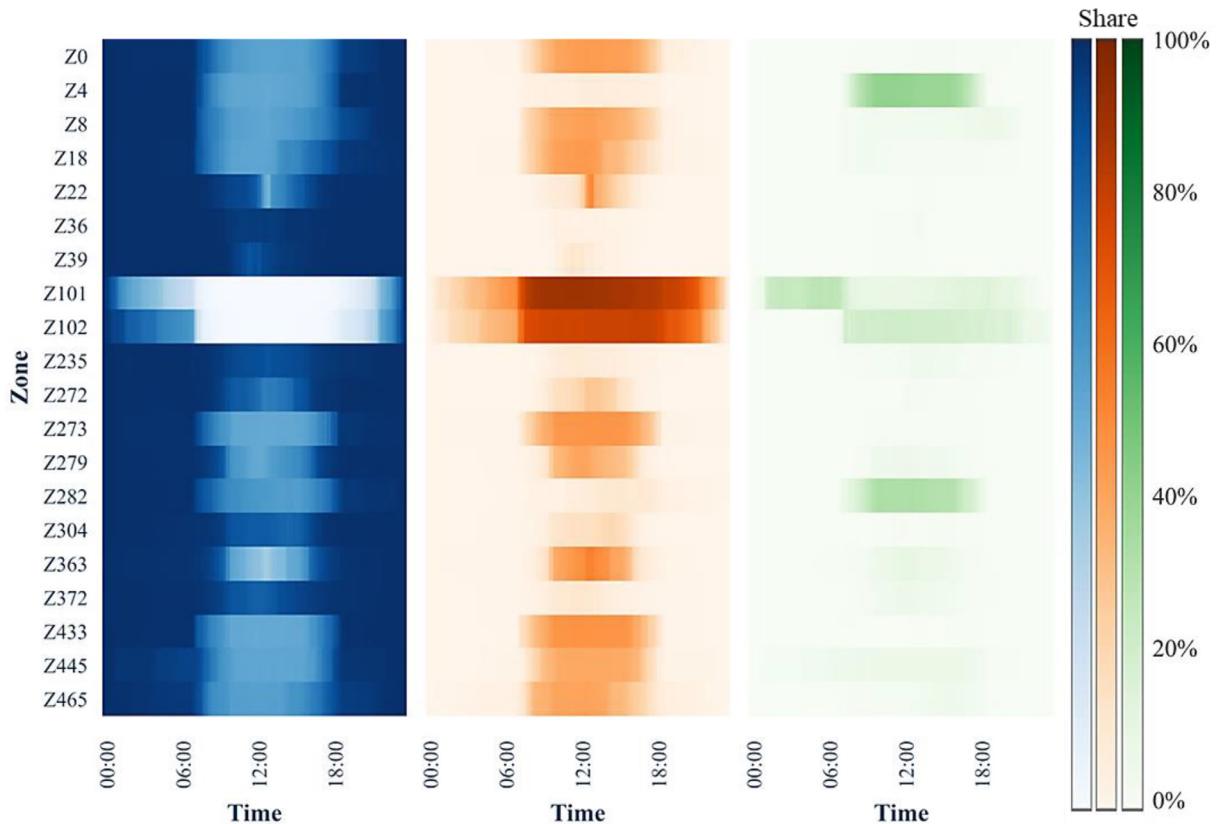


Fig. 25. Daily distribution of activity-participation for 25 randomly selected zones.

related to short duration activities. As stated earlier, these discrepancies can be potentially eliminated by the application of a more refined departure time profiles.

4.5. Travel behaviour analysis

The following section highlights the additional insight that can be drawn when aggregate ODs are converted to fully tractable activity schedules. Traditional ODs assume independency between trips therefore no assumption regarding the duration of stay between consecutive activities can be made. Nevertheless, the application of the suggested methodology allows the inference of travel patterns in great depth, something which is not possible solely on the direct analysis of aggregate OD matrices.

4.5.1. Distribution of zone activity participation

With respect to the profile of activity participation, the distribution of activities taking place during the day in the studied urban area and a sample of zones is presented in Fig. 23 and Fig. 24 respectively. As it can be observed, the patterns and the mixture of activities vary considerably across zones. Some areas present a balanced composition of activities while others present a skewed profile towards working or recreational activities. On the other hand, the aggregate diagram for the whole of the studied area presents the arguably expected pattern, with most of the out-of-home activities taking place between 08:00 to 17:00. The comparison between the aggregate and the per-zone analysis, highlights the multiple activity profiles which can arise depending on the characteristics of each zone.

On a similar stream, Fig. 25 presents the daily distribution of activities taking place at twenty-five, randomly sampled zones with an interval of 30 min. As it can be observed the activities' patterns between zones are dynamic and can vary significantly during the day. As an example, some zones (e.g. Z0, Z235, Z372) can be classified as purely residential since they are mostly occupied by their residents regardless the time of day. On the contrary zones Z101, Z448 and Z456 are mainly visited for work purposes. Nonetheless, more diverse patterns can also arise like in zone Z4 which is primarily visited for recreational and secondarily for work related activities. Another observation is that the primary activity executed in a zone can change numerous times during the day. For instance, zone Z102, is mostly occupied by its residents in the early morning,

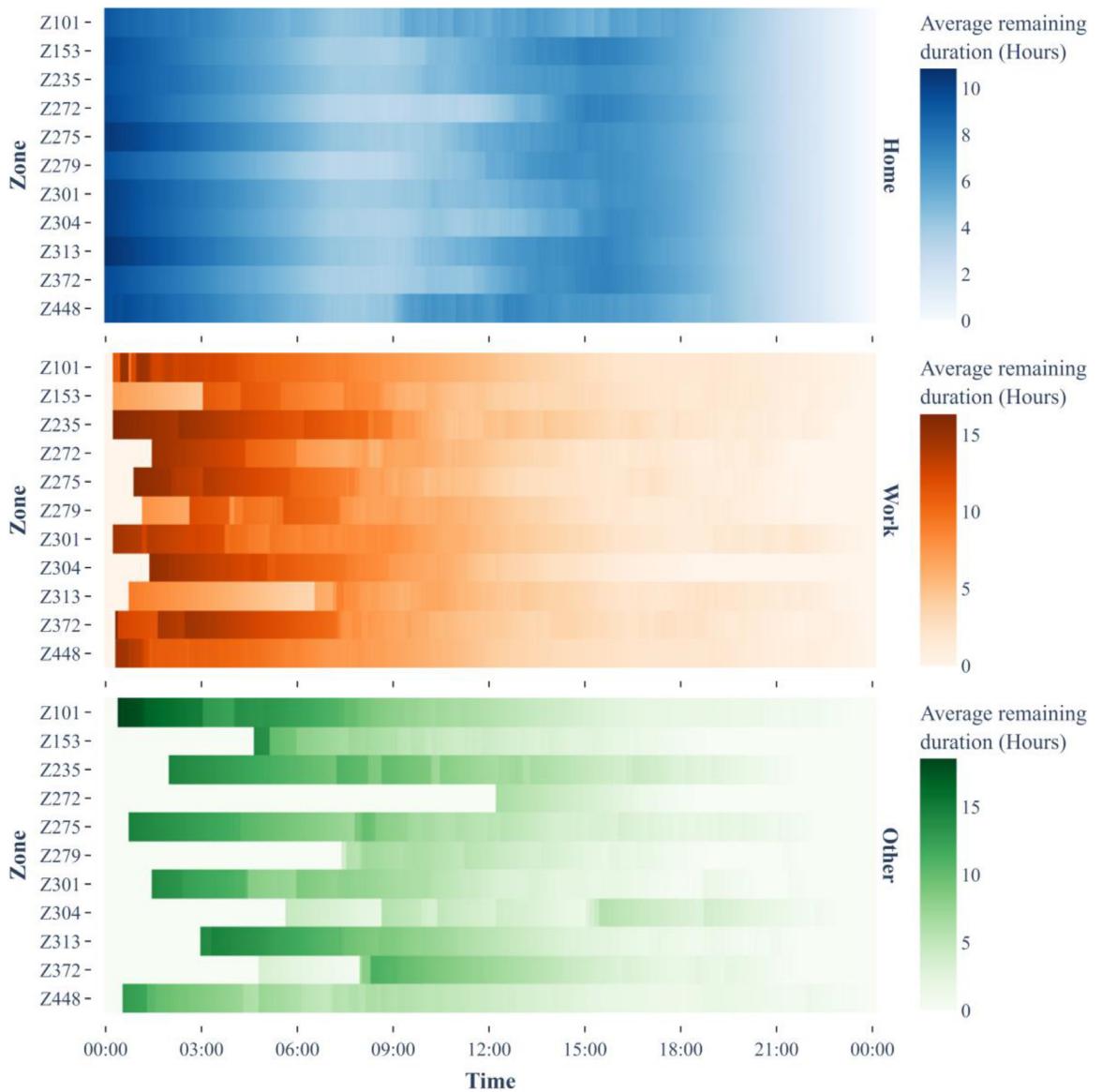


Fig. 26. Presentation of the average remaining duration of participation in activities by time of arrival and activity type.

then flooded with workers until the evening when it is visited for recreational activities before its residents return home later in the night. Such detailed information regarding the activity profile of different areas can prove particularly useful for policy-making and urban planning purposes.

4.5.2. Activity duration

Apart from the concentration of people participating in different activities, useful information can be deducted by the inclusion of the duration of the executed activities into the analysis. Fig. 26 presents the average as well as the total remaining duration of participation depending on the time of arrival and the type of activity. Prior to the interpretation of the relevant results, it must be reminded that all the studied activity schedules were completed within a single day, therefore a gradual decrease of the total and the average duration is expected. The left-hand side of the figure (Fig. 26a) depicts the average remaining time to participate in the available activities. Studying each activity type separately leads to useful insights. For instance, the average duration of stay for activity type 'Home' is long in the early morning hours as well as in the late evening when people are indeed likely to remain at their residencies for longer durations. Likewise, the methodology accurately captures the short-duration trips in the morning period (08:00 to 11:00) which can be attributed to short

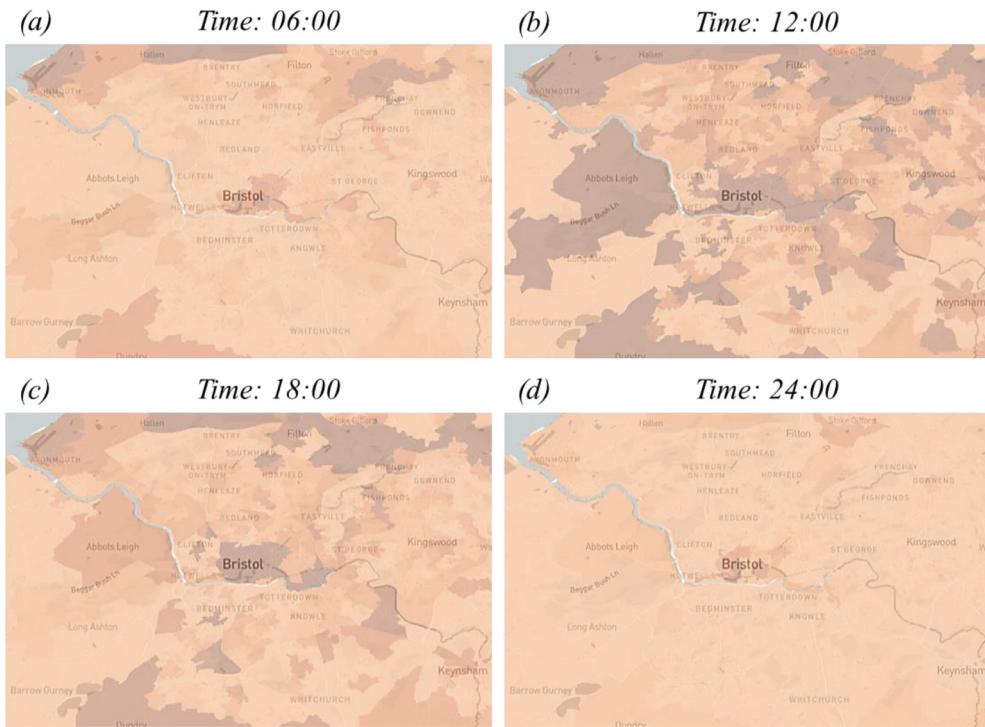


Fig. 27. Progression of participation in 'Work' type activities during a day; Darker tones indicate higher participation.

duration errands (e.g. taking children to school). On the other hand, the profile for the duration of stays for 'Work' activities differs significantly. As expected, for average duration for most of the zones is close to 8 h for arrivals to workplace between 8:00 and 10:00. Moreover, early morning workers seem to spend considerably more time at their workplace compared to those who arrive later. Finally, with regards to the 'Other' activity types, considerable spatiotemporal variation is observed. With the exception that most zones become attractive for the participation in 'Other' activity later in the day (gaps in the early hours), no other pattern seems to emerge, since the average duration of stay fluctuates both across zones and time. This element emphasises the inhomogeneous patterns observed for the activities classified as 'Other'.

4.5.3. Geospatial analysis

The last set of analyses attempts to include the geospatial factor in the process. Fig. 27 depicts the spatial distribution of people being at their workplace in the wider area of Bristol, UK, at different times of the day. As a first comment, it is obvious that the centre of Bristol is more attractive as a workplace compared to the rest of the area, regardless of the time of the day. Nonetheless, other areas in the outskirts seem to also attract a considerable share of the working population, although this share fluctuates within the day. As the day progresses, the participation in work related activities diminishes but not at the same rate for all zones, since some retain high numbers of workers even in the late evening. Including the geospatial dimension enables the identification of hotspots for different activities both in space and time and therefore allows for the evaluation of sophisticated policy scenarios.

The previously presented analysis does not aim to perform any explanatory analysis on the observed travel behaviour patterns for the studied area but rather to highlight the degree of the detail of the information that can be extracted from aggregate OD data. The direct analysis of ODs could have by no means provided enough information to complete a meticulous study of travel behaviour and urban dynamics. In contrary, the application of the suggested methodology allowed for an in-depth analysis able to unveil and properly capture the dynamic nature of urban environments.

5. Conclusions and future research

In this paper, a mechanistic modelling framework completes the preparation of detailed travel behaviour information at person level. This is achieved by the conversion of multi-period and purpose-dependant OD matrices into sets of disaggregate activity schedules, able to represent personal mobility in its full context. The suggested framework is mainly established on advanced graph-theoretical and combinatorial optimisation concepts. At first, the spatiotemporal information present in the

input OD matrices is used to create a multilayer, dynamic network (hTVG) suitable for graph-theory analysis. The application of graph-theory algorithms achieves the identification of all the plausible paths originating and ending at the same home location (i.e. tours). Then, the methodology exploits the trip-purpose information in the ODs to convert tours into activity schedules. Finally, the methodology completes the identification of the combination of activity schedules whose included trips represent the input ODs as closely as possible. The characteristics of the output can be calibrated with the provision of a distribution of the expected activity schedules' high-level characteristics (e.g. number of activities, total travel time, etc.). The validity of the methodology was evaluated based a set of multi-period and purpose dependant OD matrices including over 53,000 trips over a large urban area. The obtained results demonstrated the methodology's ability to decompose ODs accurately and efficiently into realistic activity schedules.

Despite the additional processing burden to compute tours and activity schedules out of aggregate OD matrices, the potential benefits are significant. Firstly, the disaggregate nature of the output makes it suitable for use in agent-based and microsimulation environments (e.g. MATSim, SUMO, etc.). Secondly, the level of analysis that can take place when disaggregate activity schedules are available can be significantly deeper compared to the direct analysis of aggregate ODs. Finally, the methodology can provide a mean for the creation of anonymised mobility information at person level based on sensitive urban sensing data (e.g. Mobile Phone Data, personal GPS traces).

Notwithstanding the considerable potential of the suggested methodology some areas do require further research. Firstly, the methodology should be also applied on ODs where not all trips in the ODs belong to tours. This would allow the quantification of the effect of partially observed ODs on the output. Secondly, the current framework considers each activity schedule individually, therefore it would be useful to attempt the identification of social relationships between the people completing the activity schedules (Ye et al., 2020). Finally, it should be noted that the activity schedules deriving from the presented methodology are not accompanied by sociodemographic information. However, such information is particularly important for disaggregate transport modelling (Ma and Srinivasan, 2015; McBride et al., 2018). Therefore, the possibility to couple the here proposed methodology with a population synthesis approach (Konduri et al., 2016; Ma and Srinivasan, 2016) for the assignment of demographic characteristics to the identified activity schedules should be also considered. Nonetheless, the authors suggest that the above-mentioned improvements can most likely be implemented by post-processing methodologies which will extend rather than modify the currently presented framework.

6. Appendix

Table 7
Example of input Origin-Destination matrix.

Origin Zone	Destination Zone	Purpose	Time period	Trips
Z	A	HB	AM	3
Z	A	HB	IP	1
Z	A	HB	OP	2
Z	A	NHB	IP	5
A	B	HB	IP	2
A	B	NHB	AM	3
A	B	NHB	PM	5
A	C	NHB	AM	2
A	C	NHB	PM	1
B	A	HB	AM	2
B	C	HB	PM	5
B	D	NHB	AM	3
B	D	NHB	IP	2
C	D	HB	AM	5
C	D	NHB	IP	2
C	Z	HB	OP	1
D	B	HB	PM	2
D	Z	HB	PM	3
D	Z	NHB	IP	5
D	Z	NHB	PM	2

Purpose: Home-Based (HB), Non-Home-Based (NHB).

Direction: From Home (FH), To Home (TH), Not Applicable (N/A).

Time period: Morning Peak (AM), Inter-Peak (IP), Evening Peak (PM), Off-Peak (OP).

Table 8

Example of methodology's output.

Schedule id	Activities count	Zones sequence	Trip-purposes	Activities	Time periods	Departure times
1	4	A;B;D;Z	HB;NHB;NHB;HB	H;O;O;H	IP;IP;PM;OP	11:54;12:10;16:09;22:07
2	4	A;B;D;Z	HB;NHB;NHB;HB	H;O;O;H	IP;IP;PM;OP	11:51;12:35;16:57;22:37
3	4	B;A;C;D	HB;NHB;NHB;HB	H;O;O;H	AM;AM;IP;PM	08:20;08:42;10:18;17:14
4	4	B;A;C;D	HB;NHB;NHB;HB	H;O;O;H	AM;AM;IP;PM	08:46;08:22;10:38;17:22
5	5	C;D;Z;A;B	HB;NHB;NHB;NHB;HB	H;O;O;O;H	AM;IP;IP;PM;PM	09:58;10:09;11:04;18:23;8:56
6	5	C;D;Z;A;B	HB;NHB;NHB;NHB;HB	H;O;O;O;H	AM;IP;IP;PM;PM	08:32;11:34;12:08;19:13;20:15
7	5	C;D;Z;A;B	HB;NHB;NHB;NHB;HB	H;O;O;O;H	AM;IP;IP;PM;PM	07:58;10:09;11:09;18:29;18:16
8	5	C;D;Z;A;B	HB;NHB;NHB;NHB;HB	H;O;O;O;H	AM;IP;IP;PM;PM	08:22;10:19;11:41;19:45;19:51
9	5	C;D;Z;A;B	HB;NHB;NHB;NHB;HB	H;O;O;O;H	AM;IP;IP;PM;PM	09:58;10:09;11:45;18:53;19:56
10	4	Z;A;B;D	HB;NHB;NHB;HB	H;O;O;H	AM;AM;AM;PM	08:19;09:04;09:29;18:31
11	4	Z;A;B;D	HB;NHB;NHB;HB	H;O;O;H	AM;AM;AM;PM	08:56;10:07;11:11;18:51
12	4	Z;A;B;D	HB;NHB;NHB;HB	H;O;O;H	AM;AM;AM;PM	08:29;09:54;10:19;19:31
13	3	Z;A;C	HB;NHB;HB	H;O;H	IP;PM;OP	08:55;12:45;18:42;23:21

Purpose: Home-Based (HB), Non-Home-Based (NHB).

Activities: Home (H), Other (O).

Time period: Morning Peak (AM), Inter-Peak (IP), Evening Peak (PM), Off-Peak (OP).

Table 9

Presentation of activity sequence classification.

Original activity sequence	Analysis group	Frequency	Percentage
H;W;H	H;W;H	16,534	67.078%
H;O;H	H;O;H	4883	19.810%
H;W;W;H	H;W;W;H	2012	8.163%
H;O;W;H	H;O;W;H	441	1.789%
H;W;W;W;H	Rest	282	1.144%
H;W;O;H	Rest	271	1.099%
H;W;O;W;H	Rest	75	0.304%
H;W;W;W;W;H	Rest	33	0.134%
H;O;W;W;H	Rest	28	0.114%
H;O;W;O;H	Rest	23	0.093%
H;O;O;H	Rest	19	0.077%
H;W;O;W;W;H	Rest	8	0.032%
H;O;W;W;W;H	Rest	7	0.028%
H;W;O;W;O;H	Rest	7	0.028%
H;W;W;O;H	Rest	7	0.028%
H;W;W;O;W;H	Rest	5	0.020%
H;O;W;O;W;H	Rest	4	0.016%
H;W;O;W;O;W;H	Rest	3	0.012%
H;W;W;W;W;W;H	Rest	3	0.012%
H;O;W;O;W;W;H	Rest	1	0.004%
H;W;O;O;W;H	Rest	1	0.004%
H;W;W;W;W;W;O;H	Rest	1	0.004%
H;W;W;W;W;W;W;H	Rest	1	0.004%

Table 10
Presentation of time period sequence classification.

Original time period sequence	Analysis group	Frequency
AM;PM	AM;PM	5551
AM;IP2	AM;IP2	1921
IP1;IP2	IP1;IP2	1647
IP2;PM	IP2;PM	1105
AM;IP1	AM;IP1	1102
OP1;PM	OP1;PM	1016
AM;IP3	AM;IP3	998
IP1;PM	IP1;PM	951
IP1;IP1	IP1;IP1	931
IP2;IP2	IP2;IP2	917
PM;IP3	PM;IP3	787
OP1;IP2	OP1;IP2	632
PM;PM	PM;PM	613
IP3;IP3	IP3;IP3	464
AM;AM	Rest	414
IP1;IP3	Rest	385
IP2;IP3	Rest	378
AM;PM;IP3	Rest	272
PM;OP2	Rest	268
IP3;OP2	Rest	266
OP1;IP3	Rest	248
AM;IP2;PM	Rest	234
AM;PM;PM	Rest	207
OP1;OP1	Rest	189
AM;IP1;IP2	Rest	186
AM;IP1;PM	Rest	175
IP2;OP2	Rest	162
AM;OP2	Rest	156
AM;AM;PM	Rest	134
OP1;IP1	Rest	128
IP1;IP2;PM	Rest	123
IP1;OP2	Rest	108
OP1;AM	Rest	80
IP1;IP1;IP2	Rest	77
AM;IP2;IP2	Rest	69
AM;PM;OP2	Rest	63
OP1;IP2;PM	Rest	62
AM;IP2;IP3	Rest	61
OP1;AM;PM	Rest	61
IP1;IP2;IP2	Rest	61
IP1;PM;IP3	Rest	57
AM;IP1;IP1	Rest	54
OP1;PM;IP3	Rest	47
AM;AM;IP2	Rest	47
IP1;IP2;IP3	Rest	44
OP1;PM;PM	Rest	40
IP2;PM;IP3	Rest	37
OP1;AM;IP2	Rest	35
IP1;IP1;PM	Rest	35
AM;AM;IP1	Rest	33
AM;IP1;IP2;PM	Rest	32
IP2;IP2;PM	Rest	31
IP2;PM;PM	Rest	29
AM;IP1;IP3	Rest	28
IP1;PM;PM	Rest	25
PM;IP3;OP2	Rest	23
OP1;IP1;IP2	Rest	23
AM;IP1;IP2;IP2	Rest	23
PM;IP3;IP3	Rest	22
OP1;IP1;PM	Rest	19
PM;PM;IP3	Rest	19
AM;AM;IP3	Rest	19
AM;IP3;OP2	Rest	18
OP1;AM;IP1	Rest	18
AM;IP3;IP3	Rest	17
AM;IP2;IP2;PM	Rest	17
AM;AM;IP1;IP2	Rest	15
OP1;IP2;IP3	Rest	15
AM;IP1;IP1;IP2	Rest	15

(continued on next page)

Table 10 (continued)

Original time period sequence	Analysis group	Frequency
OP1;IP2;IP2	Rest	14
AM;IP2;PM;IP3	Rest	14
AM;PM;PM;IP3	Rest	14
IP1;IP2;OP2	Rest	14
AM;AM;IP2;PM	Rest	14
IP1;IP1;IP1	Rest	13
AM;IP2;PM;PM	Rest	13
AM;PM;IP3;OP2	Rest	12
IP1;IP3;OP2	Rest	12
IP2;IP2;IP3	Rest	12
IP1;IP2;IP2;PM	Rest	12
IP2;IP2;IP2	Rest	11
IP1;PM;OP2	Rest	11
IP1;IP1;IP3	Rest	11
AM;AM;PM;IP3	Rest	10
OP1;PM;OP2	Rest	10
IP1;IP1;IP2;PM	Rest	10
IP2;IP3;IP3	Rest	10
AM;IP2;OP2	Rest	10
OP1;IP3;IP3	Rest	9
AM;AM;IP1;PM	Rest	8
OP1;IP1;IP1	Rest	8
AM;IP1;IP1;PM	Rest	8
AM;AM;PM;PM	Rest	8
OP1;AM;IP3	Rest	8
PM;PM;PM	Rest	7
IP2;IP3;OP2	Rest	7
IP1;IP2;PM;PM	Rest	7
AM;IP1;PM;IP3	Rest	7
OP1;AM;IP2;PM	Rest	7
IP1;IP1;IP2;IP2	Rest	7
AM;IP1;PM;PM	Rest	6
AM;PM;IP3;IP3	Rest	6
IP1;IP2;PM;IP3	Rest	6
OP1;AM;IP1;IP2	Rest	6
IP1;IP3;IP3	Rest	6
AM;IP1;IP1;IP2;IP2	Rest	5
AM;AM;OP2	Rest	5
OP1;OP1;IP2	Rest	5
AM;AM;IP2;IP2	Rest	4
IP1;IP1;PM;IP3	Rest	4
AM;IP1;IP2;IP3	Rest	4
AM;AM;IP1;IP2;PM	Rest	4
OP1;PM;PM;IP3	Rest	4
IP3;OP2;OP2	Rest	4
OP1;AM;AM	Rest	4
OP1;IP1;IP2;IP2	Rest	4
OP1;IP1;IP2;IP3	Rest	4
AM;IP2;IP2;IP3	Rest	4
OP1;IP2;IP2;PM	Rest	4
AM;IP2;PM;OP2	Rest	3
AM;PM;PM;PM	Rest	3
AM;IP3;IP3;OP2	Rest	3
AM;IP1;IP2;IP2;PM	Rest	3
AM;IP1;PM;PM;IP3	Rest	3
AM;IP1;IP1;IP1	Rest	3
OP1;AM;PM;PM	Rest	3
OP1;AM;AM;PM	Rest	3
OP1;AM;IP1;IP2;PM	Rest	3
OP1;AM;IP1;PM	Rest	3
OP1;AM;PM;IP3	Rest	3
IP1;IP1;OP2	Rest	3
OP1;AM;AM;IP1	Rest	3
OP1;IP1;OP2	Rest	3
OP1;IP2;OP2	Rest	3
OP1;IP3;OP2	Rest	3
IP1;IP2;IP2;IP3	Rest	3
IP2;IP2;OP2	Rest	3
IP2;IP2;PM;IP3	Rest	3

(continued on next page)

Table 10 (continued)

Original time period sequence	Analysis group	Frequency
AM;AM;IP2;PM;PM	Rest	2
AM;AM;IP1;IP1;IP2	Rest	2
IP1;IP2;PM;OP2	Rest	2
IP2;PM;IP3;IP3	Rest	2
IP2;IP2;IP3;OP2	Rest	2
IP2;IP2;IP2;PM;IP3	Rest	2
IP1;PM;PM;IP3	Rest	2
IP1;IP2;IP3;IP3	Rest	2
IP1;IP1;IP1;IP2	Rest	2
IP2;OP2;OP2	Rest	2
IP2;PM;IP3;OP2	Rest	2
AM;OP2;OP2	Rest	2
IP3;IP3;OP2	Rest	2
OP1;OP1;PM	Rest	2
PM;OP2;OP2	Rest	2
OP1;IP2;PM;OP2	Rest	2
OP1;IP1;IP2;PM	Rest	2
OP1;AM;OP2	Rest	2
OP1;AM;IP2;IP2	Rest	2
AM;PM;PM;OP2	Rest	2
OP1;PM;IP3;IP3	Rest	2
AM;IP1;IP1;IP2;PM	Rest	2
AM;IP2;IP2;IP2	Rest	2
AM;IP1;IP2;OP2	Rest	2
AM;IP1;IP1;IP3	Rest	2
AM;IP1;PM;OP2	Rest	2
AM;IP1;IP2;PM;IP3	Rest	2
AM;IP1;IP2;PM;PM;IP3	Rest	2
PM;PM;IP3;OP2	Rest	1
PM;PM;OP2	Rest	1
OP1;PM;PM;OP2	Rest	1
AM;IP1;IP1;IP1;IP2	Rest	1
OP1;IP2;IP3;OP2	Rest	1
OP1;PM;PM;IP3;IP3	Rest	1
AM;IP3;IP3;IP3	Rest	1
AM;PM;PM;PM;IP3	Rest	1
IP1;IP1;IP1;OP2	Rest	1
IP1;IP1;IP1;PM	Rest	1
IP1;IP1;IP2;IP2;PM	Rest	1
IP1;IP1;IP2;IP2;PM;IP3	Rest	1
IP1;IP1;IP2;IP2;PM;PM	Rest	1
IP1;IP1;IP2;IP3	Rest	1
IP1;IP1;IP2;PM;PM	Rest	1
IP1;IP1;PM;OP2	Rest	1
IP1;IP1;PM;PM	Rest	1
IP1;IP1;PM;PM;OP2	Rest	1
IP1;IP2;IP2;IP2	Rest	1
IP1;IP2;IP2;OP2	Rest	1
IP1;IP2;IP2;PM;IP3	Rest	1
IP1;IP2;IP2;PM;PM	Rest	1
IP1;IP2;IP2;PM;PM;PM	Rest	1
IP1;IP2;IP3;OP2	Rest	1
AM;IP2;PM;PM;IP3	Rest	1
OP1;OP1;OP1;AM;IP1	Rest	1
AM;IP2;IP3;IP3	Rest	1
AM;AM;PM;OP2	Rest	1
AM;AM;PM;IP3;IP3	Rest	1
AM;IP1;IP1;IP2;IP3	Rest	1
AM;IP1;IP1;PM;PM	Rest	1
AM;IP1;IP2;IP3;IP3	Rest	1
AM;AM;IP2;IP3	Rest	1
AM;AM;IP2;IP2;PM	Rest	1
AM;IP1;IP2;PM;OP2	Rest	1
AM;IP1;IP2;PM;PM	Rest	1
AM;AM;IP1;IP3	Rest	1
AM;AM;IP1;IP2;IP2	Rest	1
AM;IP1;IP3;OP2	Rest	1
AM;AM;IP1;IP1;OP2	Rest	1

(continued on next page)

Table 10 (continued)

Original time period sequence	Analysis group	Frequency
AM;AM;IP1;IP1;IP2;IP2;IP2	Rest	1
AM;IP1;OP2;OP2	Rest	1
AM;AM;AM;PM	Rest	1
AM;AM;AM;IP2	Rest	1
AM;AM;AM;IP1	Rest	1
AM;IP1;PM;PM;IP3;OP2	Rest	1
AM;IP2;IP2;IP2;PM	Rest	1
AM;IP2;IP2;IP3;IP3	Rest	1
AM;IP2;IP2;OP2	Rest	1
AM;IP2;IP2;PM;OP2	Rest	1
IP1;IP2;PM;IP3;OP2	Rest	1
IP1;IP2;PM;PM;IP3	Rest	1
IP1;PM;IP3;IP3	Rest	1
OP1;AM;IP2;IP3	Rest	1
OP1;AM;IP2;PM;IP3	Rest	1
OP1;AM;IP3;IP3	Rest	1
OP1;AM;PM;OP2	Rest	1
OP1;IP1;IP1;IP2	Rest	1
OP1;IP1;IP1;IP2;PM	Rest	1
OP1;IP1;IP1;PM	Rest	1
OP1;IP1;IP2;OP2	Rest	1
OP1;IP1;IP2;PM;PM	Rest	1
OP1;IP1;IP3	Rest	1
OP1;IP1;PM;IP3	Rest	1
OP1;IP2;IP2;IP3	Rest	1
AM;IP1;IP1;IP1;IP2;PM;PM	Rest	1
OP1;IP2;PM;IP3	Rest	1
OP1;IP2;PM;PM	Rest	1
OP1;OP1;AM	Rest	1
OP1;OP1;AM;AM;OP2;OP2	Rest	1
OP1;OP1;IP2;PM	Rest	1
OP1;OP1;OP1	Rest	1
OP1;AM;IP1;IP1;IP3	Rest	1
OP1;AM;IP1;IP1;IP2	Rest	1
IP2;PM;PM;IP3	Rest	1
IP1;PM;PM;OP2	Rest	1
IP2;IP2;IP2;PM	Rest	1
IP2;IP2;PM;OP2	Rest	1
IP2;IP2;PM;PM	Rest	1
IP2;IP3;IP3;OP2	Rest	1
OP1;AM;AM;IP2;PM	Rest	1
IP2;PM;PM;PM;IP3	Rest	1
IP3;IP3;IP3	Rest	1
OP1;AM;AM;IP1;IP1	Rest	1
OP1;AM;AM;IP1;IP2	Rest	1
OP1;AM;AM;IP1;IP3	Rest	1
OP1;AM;AM;IP1;PM	Rest	1
OP1;AM;AM;IP2	Rest	1
IP1;IP1;IP2;IP2;IP2	Rest	1

CRediT authorship contribution statement

Haris Ballis: Conceptualization, Methodology, Visualization, Software, Investigation, Writing - original draft. **Loukas Dimitriou:** Supervision, Conceptualization, Methodology, Writing - original draft.

References

- Allahviranloo, M., Recker, W., 2013. Daily activity pattern recognition by using support vector machines with multiple classes. *Transp. Res. Part B Methodol.* 58, 16–43. <https://doi.org/10.1016/j.trb.2013.09.008>.
- Anda, C., Fourie, P., Erath, A., 2016. *Transport Modelling in the Age of Big Data*. Singapore - ETH Cent. Futur. Cities Lab Work Report.
- Antoniou, C., Dimitriou, L., Pereira, F., 2019. Mobility patterns, Big Data and Transport analytics : Tools and Applications For Modeling. Elsevier.
- Avegliano, P., Sichman, J.S., 2019. Using Surrogate Models to Calibrate Agent-based Model Parameters Under Data Scarcity, in: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp. 1781–1783.
- Axhausen, K.W., 2007. Concepts of Travel Behaviour Research, in: *Threats from Car Traffic to the Quality of Urban Life*. Emerald Group Publishing Limited, pp. 165–185.
- Axhausen, K.W., Horni, A., Nagel, K., Axhausen, K.W., 2016. The Multi-Agent Transport Simulation MATSim. The Multi-Agent Transport Simulation MATSim. Ubiquity Press <https://doi.org/10.5334/baw>.
- Bachir, D., Khodabandehlou, G., Gauthier, V., El Yacoubi, M., Puchinger, J., 2019. Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transp. Res. Part C Emerg. Technol.* 101, 254–275. <https://doi.org/10.1016/j.trc.2019.02.013>.
- Bakker, M.M., Alam, S.J., van Dijk, J., Rounsevell, M.D.A., 2014. Land-use change arising from rural land exchange: an agent-based simulation model. *Landsc. Ecol.* 30, 273–286. <https://doi.org/10.1007/s10980-014-0116-x>.
- Ballis, H., Dimitriou, L., 2019. Optimal Population of Trip Chains Synthesis from Multi-Period Origin-Destination Matrices, in: *Proceedings of Transportation Research Board 98th Annual Meeting*, Washington D.C.
- Barabási, A.-L., 2016. *Network Science*. Cambridge university press.
- Barthelemy, J., Toint, P.L., 2013. Synthetic Population Generation Without a Sample. *Transp. Sci.* 47, 266–279. <https://doi.org/10.1287/trsc.1120.0408>.
- Ben-Akiva, M., Bottoni, J., Gao, S., Koutsopoulos, H.N., Wen, Y., 2007. Towards Disaggregate Dynamic Travel Forecasting Models. *Tsinghua Sci. Technol* 12, 115–130. [https://doi.org/10.1016/S1007-0214\(07\)70019-6](https://doi.org/10.1016/S1007-0214(07)70019-6).
- Bhat, C.R., 1996. A generalized multiple durations proportional hazard model with an application to activity behavior during the evening work-to-home commute. *Transp. Res. Part B Methodol.* 30, 465–480. [https://doi.org/10.1016/0191-2615\(96\)00007-0](https://doi.org/10.1016/0191-2615(96)00007-0).
- Bhat, C.R., Koppelman, F.S., 1999. Activity-Based Modeling of Travel Demand 35–61. https://doi.org/10.1007/978-1-4615-5203-1_3
- Bhat, C.R., Srinivasan, S., Axhausen, K.W., 2005. An analysis of multiple interepisode durations using a unifying multivariate hazard model. *Transp. Res. Part B Methodol.* 39, 797–823. <https://doi.org/10.1016/j.trb.2004.11.002>.
- Bonnel, P., Hombourger, E., Olteanu-Raimond, A.M., Smoreda, Z., 2015. *Passive Mobile Phone Dataset to Construct Origin-Destination matrix: Potentials and limitations*, in: *Transportation Research Procedia*. Elsevier, pp. 381–398.
- Bowman, J.L., 1998. *The Day Activity Schedule Approach to Travel Demand Analysis*. Metro, p. 185.
- Bowman, J.L., Ben-Akiva, M., 2000. Activity-based disaggregate travel demand model system with activity schedules. *Transp. Res. Part A Policy Pract.* 35, 1–28. [https://doi.org/10.1016/S0965-8564\(99\)00043-9](https://doi.org/10.1016/S0965-8564(99)00043-9).
- Bradley, M., Bowman, J.L., Griesenbeck, B., 2010. SACSIM: an applied activity-based model system with fine-level spatial and temporal resolution. *J. Choice Model.* 3, 5–31. [https://doi.org/10.1016/S1755-5345\(13\)70027-7](https://doi.org/10.1016/S1755-5345(13)70027-7).
- Caceres, N., Romero, L.M., Benitez, F.G., 2013. Inferring origin-destination trip matrices from aggregate volumes on groups of links: a case study using volumes inferred from mobile phone data. *J. Adv. Transp.* 47, 650–666. <https://doi.org/10.1002/atr.187>.
- Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., Ratti, C., Lorenzo, G.Di, Ferreira, J., Ratti, C., 2013. Understanding individual mobility patterns from urban sensing data - A mobile phone trace example. *Transp. Res. Part C Emerg. Technol.* 26, 301–313. <https://doi.org/10.1016/j.trc.2012.09.009>.
- Cantelmo, G., Qurashi, M., Prakash, A.A., Antoniou, C., Viti, F., 2019. Incorporating trip chaining within online demand estimation. *Transp. Res. Part B Methodol.* <https://doi.org/10.1016/j.trb.2019.05.010>.
- Casteigts, A., 2018. Finding Structure in Dynamic Networks.
- Chen, C., Bian, L., Ma, J., 2014. From traces to trajectories: how well can we guess activity locations from mobile phone traces? *Transp. Res. Part C Emerg. Technol* 46, 326–337. <https://doi.org/10.1016/j.trc.2014.07.001>.
- Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transp. Res. Part C Emerg. Technol.* <https://doi.org/10.1016/j.trc.2016.04.005>.
- Cheng, E., Grossman, J.W., Lipman, M.J., 2003. Time-stamped graphs and their associated influence digraphs. *Discret. Appl. Math.* 128, 317–335. [https://doi.org/10.1016/S0166-218X\(02\)00497-3](https://doi.org/10.1016/S0166-218X(02)00497-3).
- Cich, G., Knappen, L., Maciejewski, M., Yasar, A.U.H., Bellmans, T., Janssens, D., 2017. Modeling Demand Responsive Transport using SARL and MATSim. in: *Procedia Computer Science* 1074–1079. <https://doi.org/10.1016/j.procs.2017.05.387>.
- Çolak, S., Alexander, L.P., Alvim, B.G., Mehndiratta, S.R., Gonzalez, M.C., 2015. Analyzing Cell Phone Location Data for Urban Travel. *Transp. Res. Rec. J. Transp. Res. Board* 2526, 126–135. <https://doi.org/10.3141/2526-14>.
- Department for Transport, 2017. *National Travel Survey: england 2016. National Travel Survey*.
- Djavadian, S., Chow, J.Y.J., 2017. An agent-based day-to-day adjustment process for modeling 'Mobility as a Service' with a two-sided flexible transport market. *Transp. Res. Part B Methodol.* 104, 36–57. <https://doi.org/10.1016/j.trb.2017.06.015>.
- E. Ramadan, O., Sisiopiku, V., P., 2019. A Critical Review on Population Synthesis for Activity- and Agent-Based Transportation Models, in: *transportation [Working Title]*. IntechOpen.. <https://doi.org/10.5772/intechopen.86307>.
- Eagle, N., Pentland, A.S., 2009. Eigenbehaviors: identifying structure in routine. *Behav. Ecol. Sociobiol.* 63, 1057–1066. <https://doi.org/10.1007/s00265-009-0739-0>.
- Ebadí, N., Kang, J.E., Hasan, S., 2017. Constructing activity-mobility trajectories of college students based on smart card transaction data. *Int. J. Transp. Sci. Technol.* 6, 316–329. <https://doi.org/10.1016/j.ijtst.2017.08.003>.
- Ettema, D., Borgers, A., Timmermans, H., 1995. Competing risk hazard model of activity choice, timing, sequencing, and duration. *Transp. Res. Rec.* 1493, 101–109.
- Ferreira, A., 2004. Building a reference combinatorial model for MANETs. *IEEE Netw* 18, 24–29. <https://doi.org/10.1109/MNET.2004.1337732>.
- Flötteröd, G., Bierlaire, M., Nagel, K., 2011. Bayesian demand calibration for dynamic traffic simulations. *Transp. Sci.* 45, 541–561. <https://doi.org/10.1287/trsc.1100.0367>.
- Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L., 2008. Understanding individual human mobility patterns 453, 779–782. <https://doi.org/10.1038/nature06958>.
- Goodwin, P.B., 1981. The usefulness of travel budgets. *Transp. Res. Part A Gen.* 15, 97–106. [https://doi.org/10.1016/0191-2607\(83\)90019-5](https://doi.org/10.1016/0191-2607(83)90019-5).
- Goulet Langlois, G., Koutsopoulos, H.N., Zhao, J., 2016. Inferring patterns in the multi-week activity sequences of public transport users. *Transp. Res. Part C Emerg. Technol.* 64, 1–16. <https://doi.org/10.1016/j.trc.2015.12.012>.
- Goulias, K.G., Kitamura, R., 1991. Recursive Model System for Trip Generation and Trip Chaining. *Transp. Res. Rec* 59–66.
- Han, G., Sohn, K., 2016. Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model. *Transp. Res. Part B Methodol.* 83, 121–135. <https://doi.org/10.1016/j.trb.2015.11.015>.

- Hart, W.E., Laird, C.D., Watson, J.-P., Woodruff, D.L., Hackebeil, G.A., Nicholson, B.L., Siirola, J.D., 2017. Pyomo – Optimization Modeling in Python. <https://doi.org/10.1007/978-3-319-58821-6>
- IBM, 2020. CPLEX Optimizer | IBM [WWW Document]. URL <https://www.ibm.com/analytics/cplex-optimizer> (accessed 7.28.18).
- Iqbal, M.S., Choudhury, C.F., Wang, P., Gonzalez, M.C., 2014. Development of origin-destination matrices using mobile phone call data. *Transp. Res. Part C Emerg. Technol.* 40, 63–74. <https://doi.org/10.1016/j.trc.2014.01.002>.
- Jiang, S., Ferreira, J., González, M.C., 2012. Clustering daily patterns of human activities in the city, in: *Data Min Knowl Discov* 478–510. <https://doi.org/10.1007/s10618-012-0264-z>.
- Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S., González, M.C., 2016. The TimeGeo modeling framework for urban motility without travel surveys, in: *proceedings of the National Academy of Sciences of the United States of America*. pp. E5370–E5378. <https://doi.org/10.1073/pnas.1524261113>
- Joh, C.H., Arentze, T., Hofman, F., Timmermans, H., 2002. Activity pattern similarity: a multidimensional sequence alignment method. *Transp. Res. Part B Methodol.* 36, 385–403. [https://doi.org/10.1016/S0191-2615\(01\)00009-1](https://doi.org/10.1016/S0191-2615(01)00009-1).
- Jun, C., Donghyun, Y., 2013. Estimating smart card commuters origin-destination distribution based on APTS data. *J. Transp. Syst. Eng. Inf. Technol.* 13, 47–53. [https://doi.org/10.1016/S1570-6672\(13\)60116-6](https://doi.org/10.1016/S1570-6672(13)60116-6).
- Klotz, E., Newman, A.M., 2013. Practical guidelines for solving difficult mixed integer linear programs. *Surv. Oper. Res. Manag. Sci.* <https://doi.org/10.1016/j.sorms.2012.12.001>.
- Konduri, K.C., You, D., Garikapati, V.M., Pendyala, R.M., 2016. Enhanced synthetic population generator that accommodates control variables at multiple geographic resolutions. *Transp. Res.* 2563, 40–50. <https://doi.org/10.3141/2563-08>.
- Kostakos, V., 2009. Temporal graphs. *Phys. A Stat. Mech. its Appl.* 388, 1007–1023. <https://doi.org/10.1016/j.physa.2008.11.021>.
- Kumar, R., Calders, T., 2018. 2SCENT: an Efficient algorithm for enumerating all simple temporal cycles. *Proc. VLDB Endow.* 11, 1441–1453. <https://doi.org/10.14778/3236187.3236197>.
- Lee, Y., Hickman, M., Washington, S., 2007. Household type and structure, time-use pattern, and trip-chaining behavior. *Transp. Res. Part A Policy Pract.* 41, 1004–1020. <https://doi.org/10.1016/j.tra.2007.06.007>.
- Lin, X., Sun, W., Veeraraghavan, M., Hu, W., 2016. Time-shifted multilayer graph: a routing framework for bulk data transfer in optical circuit-switched networks with assistive storage. *J. Opt. Commun. Netw.* 8, 162–174. <https://doi.org/10.1364/JOCN.8.000162>.
- Lin, Z., Yin, M., Feygin, S., Sheehan, M., Paiement, J.-F., Pozdnoukhov, A., 2017. Deep Generative Models of Urban Mobility. *ACM SIGKDD Conf.* 9 <https://doi.org/10.475/123>.
- Lindveld, C.D.R., 2003. Dynamic O-D matrix estimation: a behavioural approach.
- Liu, F., Janssens, D., Cui, J., Wang, Y., Wets, G., Cools, M., 2014. Building a validation measure for activity-based transportation models based on mobile phone data. *Expert Syst. Appl.* 41, 6174–6189. <https://doi.org/10.1016/j.eswa.2014.03.054>.
- Liu, F., Janssens, D., Cui, J., Wets, G., Cools, M., 2015. Characterizing activity sequences using profile Hidden Markov Models. *Expert Syst. Appl.* 42, 5705–5722. <https://doi.org/10.1016/j.eswa.2015.02.057>.
- Liu, F., Janssens, D., Wets, G., Cools, M., 2013. Annotating mobile phone location data with activity purposes using machine learning algorithms. *Expert Syst. Appl.* 40, 3299–3311. <https://doi.org/10.1016/j.eswa.2012.12.100>.
- Ma, J., Li, H., Yuan, F., Bauer, T., 2013. Deriving Operational Origin-Destination Matrices From Large Scale Mobile Phone Data. *Int. J. Transp. Sci. Technol.* 2, 183–204. <https://doi.org/10.1260/2046-0430.2.3.183>.
- Ma, L., Srinivasan, S., 2016. An empirical assessment of factors affecting the accuracy of target-year synthetic populations. *Transp. Res. Part A Policy Pract.* 85, 247–264. <https://doi.org/10.1016/j.tra.2016.01.016>.
- Ma, L., Srinivasan, S., 2015. Synthetic population generation with multilevel controls: a fitness-based synthesis approach and validations. *Comput. Civ. Infrastruct. Eng.* 30, 135–150. <https://doi.org/10.1111/mice.12085>.
- McBride, E.C., Davis, A.W., Goulias, K.G., 2018. A Spatial Latent Profile Analysis to Classify Land Uses for Population Synthesis Methods in Travel Demand Forecasting. *Transp. Res. Rec.* 2672, 158–170. <https://doi.org/10.1177/036198118799168>.
- McGuckin, N., Murakami, E., 1999. Examining Trip-Chaining Behavior: comparison of Travel by Men and Women. *Transp. Res. Rec. J. Transp. Res. Board* 1693, 79–85. <https://doi.org/10.3141/1693-12>.
- Mcnally, M.G., Rindt, C., 2008. The Activity-Based Approach. In: Hensher, D.A., Button, K. (Eds.), *Handbook of Transport Modelling*. Emerald Group Publishing Limited.
- Morimura, T., Osogami, T., 2013. Solving inverse problem of Markov chain with partial observations.
- Ni, B., Shen, Q., Xu, J., Qu, H., 2017. Spatio-temporal flow maps for visualizing movement and contact patterns. *Vis. Informatics* 1, 57–64. <https://doi.org/10.1016/j.visinf.2017.01.007>.
- Nie, Y., Zhang, H.M., Recker, W.W., 2005. Inferring origin-destination trip matrices with a decoupled GLS path flow estimator. *Transp. Res. Part B Methodol.* 39, 497–518. <https://doi.org/10.1016/j.trb.2004.07.002>.
- Nurul Habib, K.M., 2011. A random utility maximization (RUM) based dynamic activity scheduling model: application in weekend activity scheduling. *Transportation (Amst)* 38, 123–151. <https://doi.org/10.1007/s11116-010-9294-9>.
- Ortúzar, J., de, D., Willumsen, L.G., 2011. *Modelling Transport*. Modelling Transport, 4th ed Wiley-Blackwell <https://doi.org/10.1002/9781119993308>.
- Ozimek, A., Miles, D., 2011. Stata utilities for geocoding and generating travel time and travel distance information. *Stata J* 11, 106–119. <https://doi.org/10.1177/1536867x1101100107>.
- Pan, C., Lu, J., Di, S., Ran, B., 2006. Cellular-based data-extracting method for trip distribution. in: *Transportation Research Record* 33–39. <https://doi.org/10.3141/1945-04>.
- Pappalardo, L., Simini, F., 2018. Data-driven generation of spatio-temporal routines in human mobility. *Data Min. Knowl. Discov.* 32, 787–829. <https://doi.org/10.1007/s10618-017-0548-4>.
- Pendyala, R.M., Goulias, K.G., 2002. Time use and activity perspectives in travel behavior research. *Transportation (Amst)* 29, 1–4.
- Phan, D., Xiao, L., Yeh, R., Hanrahan, P., Winograd, T., 2005. Flow map layout. *Proc. - IEEE Symp. Inf. Vis. INFO VIS* 219–224. <https://doi.org/10.1109/INFVIS.2005.1532150>.
- Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibusaki, R., Ratti, C., 2010. Activity-aware map: identifying human daily activity pattern using mobile phone data, in: *lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 14–25. https://doi.org/10.1007/978-3-642-14715-9_3
- Pinjari, A.R., Bhat, C.R., 2011. Activity-based Travel Demand Analysis. In: de Palma, A., Lindsey, R., Quinet, E. (Eds.), *A Handbook of Transport Economics*. Edward Elgar Publishing Ltd. pp. 213–248 <https://doi.org/10.4337/9780857930873.00017>.
- Primerano, F., Taylor, M.A.P., Pitakringkarn, L., Tisato, P., 2008. Defining and understanding trip chaining behaviour. *Transportation (Amst)*. <https://doi.org/10.1007/s11116-007-9134-8>.
- Raux, C., Ma, T.Y., Cornelis, E., 2016. Variability in daily activity-travel patterns: the case of a one-week travel diary. *Eur. Transp. Res. Rev.* 8. <https://doi.org/10.1007/s12544-016-0213-9>.
- Recker, W.W., 2001. A bridge between travel demand modeling and activity-based travel analysis. *Transp. Res. Part B Methodol.* 35, 481–506. [https://doi.org/10.1016/S0191-2615\(00\)00006-0](https://doi.org/10.1016/S0191-2615(00)00006-0).
- Redondo, J.L., Pelegrin, B., Fernandez, P., Garcia, I., Ortigosa, P.M., 2011. Finding multiple global optima for unconstrained discrete location problems. *Optim. Methods Softw.* 26, 207–224. <https://doi.org/10.1080/10556780903567760>.
- Ronald, N., Thompson, R., Winter, S., 2015. Simulating Demand-responsive Transportation: a Review of Agent-based Approaches. *Transp. Rev.* 35, 404–421. <https://doi.org/10.1080/01441647.2015.1017749>.
- Saadi, I., Mustafa, A., Teller, J., Cools, M., 2016. Forecasting travel behavior using Markov Chains-based approaches. *Transp. Res. Part C Emerg. Technol.* 69, 402–417. <https://doi.org/10.1016/j.trc.2016.06.020>.

- Santoro, N., Quattrociocchi, W., Flocchini, P., Casteigts, A., Amblard, F., 2011. Time-varying graphs and social network analysis: temporal indicators and metrics, in: AISB 2011. Social Networks and Multiagent Systems 33–38.
- Schneider, C.M., Belik, V., Couronné, T., Smoreda, Z., González, M.C., 2013. Unravelling daily human mobility motifs. *J. R. Soc. Interface* 10, 20130246. <https://doi.org/10.1098/rsif.2013.0246>.
- Schneider, F., Ton, D., Zomer, L.-B., Daamen, W., Duives, D., Hoogendoorn-Lanser, S., Hoogendoorn, S., 2020. Trip chain complexity: a comparison among latent classes of daily mobility patterns. *Transportation (Amst)* 1–23. <https://doi.org/10.1007/s11116-020-10084-1>.
- Schoenfelder, S., Axhausen, K.W., 2001. Analysing the rhythms of travel using survival analysis, in: transportation Research Board (TRB) Annual Meeting. <https://doi.org/10.3929/ETHZ-A-004241369>
- Sedgewick, R., 2001. Algorithms in C. Addison-Wesley Professional [https://doi.org/10.1016/0965-9978\(92\)90046-i](https://doi.org/10.1016/0965-9978(92)90046-i).
- Song, C., Qu, Z., Blumm, N., Barabási, A.L., 2010. Limits of predictability in human mobility. *Science* 80–. (327), 1018–1021. <https://doi.org/10.1126/science.117170>.
- Srikrishnan, V., Keller, K., 2018. How much data are needed to calibrate and test agent-based models?
- Tesselkin, A., Khabarov, V., 2017. Estimation of Origin-Destination Matrices Based on Markov Chains. *Procedia Eng.* 178, 107–116. <https://doi.org/10.1016/j.proeng.2017.01.071>.
- Thill, J.-C., Thomas, I., 1987. Toward Conceptualizing Trip-Chaining Behavior: a Review. *Geogr. Anal.* 19, 1–17. <https://doi.org/10.1111/j.1538-4632.1987.tb00110.x>.
- Tolouei, R., Psarras, S., Prince, R., 2017. Origin-Destination Trip Matrix Development: Conventional Methods Versus Mobile Phone Data, in: *Transportation Research Procedia*. Elsevier, pp. 39–52.
- Toole, J.L., Colak, S., Sturt, B., Alexander, L.P., Evsukoff, A., Gonzalez, M.C., 2015. The path most traveled: travel demand estimation using big data resources. *Transp. Res. Part C Emerg. Technol.* 58, 162–177. <https://doi.org/10.1016/j.trc.2015.04.022>.
- Vlahogianni, E.I., Park, B.B., van Lint, J.W.C.W.C., 2015. Big data in transportation and traffic engineering. *Transp. Res. Part C Emerg. Technol.* 58, 161. <https://doi.org/10.1016/j.trc.2015.08.006>.
- Von Landesberger, T., Brodkorb, F., Roskosch, P., Andrienko, N., Andrienko, G., Kerren, A., 2016. MobilityGraphs: visual Analysis of Mass Mobility Dynamics via Spatio-Temporal Graphs and Clustering. *IEEE Trans. Vis. Comput. Graph.* 22, 11–20. <https://doi.org/10.1109/TVCG.2015.246811>.
- Wang, W., Attanucci, J.P., Wilson, N.H.M.M., 2011. Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems. *J. Public Transp.* 14, 131–150. <https://doi.org/10.5038/2375-0901.14.4.7>.
- Wang, Y., Yuan, Y., Ma, Y., Wang, G., 2019. Time-Dependent Graphs: definitions, Applications, and Algorithms. *Data Sci. Eng.* <https://doi.org/10.1007/s41019-019-00105-0>.
- Wehmuth, K., Viviani, A., Fleury, E., 2015. A Unifying Model For Representing Time-Varying graphs, in: *Proceedings of the 2015 IEEE International Conference On Data Science and Advanced Analytics. DSAA 2015*. Institute of Electrical and Electronics Engineers Inc.
- Wilson, W.C., 1998. Activity pattern analysis by means of sequence-alignment methods. *Environ. Plan. A* 30, 1017–1038. <https://doi.org/10.1068/a301017>.
- Wood, J., Dykes, J., Slingsby, A., 2010. Visualisation of Origins, Destinations and Flows with OD Maps. *Cartogr. J* 47, 117–129. <https://doi.org/10.1179/000870410x12658023467367>.
- Ye, P., Zhu, F., Sabri, S., Wang, F.Y., 2020. Consistent Population Synthesis with Multi-Social Relationships Based on Tensor Decomposition. *IEEE Trans. Intell. Transp. Syst.* 21, 2180–2189. <https://doi.org/10.1109/TITS.2019.2916867>.
- Yue, Y., Lan, T., Yeh, A.G.O., Li, Q.Q., 2014. Zooming into individuals to understand the collective: a review of trajectory-based travel behaviour studies. *Travel Behav. Soc.* 1, 69–78. <https://doi.org/10.1016/j.tbs.2013.12.002>.
- Zhang, W., Thill, J.C., 2017. Detecting and visualizing cohesive activity-travel patterns: a network analysis approach. *Comput. Environ. Urban Syst.* 66, 117–129. <https://doi.org/10.1016/j.compenvurbsys.2017.08.004>.
- Zhao, J., Rahbee, A., Wilson, N.H.M., 2007. Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Comput. Civ. Infrastruct. Eng.* 22, 376–387. <https://doi.org/10.1111/j.1467-8667.2007.00494.x>.
- Zilske, M., Nagel, K., 2015. A simulation-based approach for constructing all-day travel chains from mobile phone data. in: *Procedia Computer Science* 468–475. <https://doi.org/10.1016/j.procs.2015.05.017>.