Preparation of agent-based modelling input from aggregate 4-step model

Haris Ballis, Ph.D.[1], Loukas Dimitriou, Ph.D.*[2]

[1] Post-doc Researcher, LαB for Transport Engineering Dept. of Civil and Environmental Engineering University of Cyprus

[2] Head, LαB for Transport Engineering, Dept. of Civil and Environmental Engineering, University of Cyprus

SHORT SUMMARY

Agent-based and activity-based modelling are considered as the most suitable means to model the emerging mobility environment (New Mobility Services, Mobility as a Service, ride-hailing, etc.) but the lack of experience on the development of such models as well as their considerable requirements in fine-grained input have hindered their wider adoption. For those reasons, and despite their shortcoming, aggregate 4-step models still constitute the most wide-spread approach in the field of travel demand modelling. This study evaluates the possibility to exploit the outputs of existing 4-step transport models for the preparation of input suitable for agent- and activity-based modelling. In particular, the typical outputs of 4-step models, namely, multi-period Origin-Destination matrices (ODs) and traffic assignment results (i.e. link counts) are transformed into home-based trip-chains (i.e. tours). The framework is evaluated on the 4-step model of Larnaka, Cyprus.

Keywords: OD Matrices; Disaggregation; Large-scale Combinatorial Optimisation

1. Introduction

Shortcomings of the macroscopic transport modelling approaches

The typical approach to model mobility encompasses the use of aggregate transport models. Traditional aggregate transport models (such as the widely used 4-step model) have proven invaluable tools towards the understanding of transportation trends and have tremendously supported crucial strategic decisions. However, the future mobility challenges are calling for the development of disaggregate modelling paradigms, better suited to capture individuals' travel behaviour and complex multiagent systems (Goulias 2009). This statement has been supported by many researchers who have emphasised the need of incorporating user-centricity in the core of transport modelling for mitigation against future mobility challenges (Hilgert et al. 2016; Ben-Akiva et al. 2007). However, the introduction of user-centricity in typical aggregate (i.e. 4-step) transport models can prove troublesome since their aggregate nature hinders the representation of travel behaviour at person-level (Mladenovic and Trifunovic 2014). Traditional aggregate transport models were not designed with the aim to evaluate the travellers' reaction to personalised transport services such as NMS (Pinjari and Bhat 2011).

Instead, aggregate transport models typically apply analytical relationships, based on closed form mathematical equations to estimate total travel demand between pairs of zones (Ben-Akiva et al. 2007). Attempting to include the individual characteristics of each user to the modelling framework of 4-step models would require the impractical introduction of a unique user class for each traveller. On the other hand, disaggregate paradigms are considerably more capable of incorporating user-centricity into transport and travel behaviour modelling since different agents can be natively modelled.

Apart from the negative implications of using typical aggregate models for the prediction of travel behaviour under the emerging flexible and on-demand transport services, difficulties also arise for the modelling of the operational side of transport (Segui-Gasco et al. 2019; Horn 2002). The future of transport promises seamless journeys which can possibly require multiple complementing modes and cooperating transport operators (Kamargianni et al. 2015; Smith, Sochor, and Karlsson 2018). Such a complex system will not operate efficiently support from sophisticated modelling tools able to incorporate highly dynamic networks, inhomogeneous fleets, and continuously changing travel behaviours (Cich et al. 2017). In addition, these modelling tools should be also able to predict future conditions and suggest appropriate mitigation strategies in case of emergencies or unforeseen events. From the transport supply perspective aggregate transport models are in a weak position (compared to disaggregate models) to effectively model the dynamic and multipart operations manifesting in the emerging transportation landscape (Cich et al. 2016).

ODs to disaggregate tours

The increasing requirements for high precision, disaggregate mobility information, in conjunction with the data-privacy regulations (i.e. GDPR-EU, APPI-Japan, etc.) which promote the aggregated publishing of information has led researchers to experiment with data disaggregation methodologies. Recently, Huber and Lißner (2019) utilised aggregate cycling data obtained from the Strava app to synthesise disaggregate mobility data. Their approach applies a double-constrained routing algorithm on aggregate OD cycling demand to derive single bicycle routes. However, their model does not aim at the reproduction of the cycling travel demand through individual cycling traces but rather on the development of a bicycle route choice model based on the OD information. The possibility of synthesising travel demand based on aggregated data from Telecommunication Service Providers (TSPs) has been recently evaluated by Anda et al. (2020). Their Markov-based approach allows the synthesis of realistic daily tours using aggregate joint distributions (histograms) which can be provided by TSPs since they are considerably less likely to raise data-privacy concerns. Multiple different model architectures were evaluated over a large dataset of 1 million synthetic travellers and resulted in remarkably high accuracy ($\geq 95\%$) in terms of replicating the observed travel patterns. A potential drawback of the methodology is the reliance on multiple and very detailed hourly distributions at zonal level (e.g. duration of stay time in a zone by hour, number of people transitioning to a previously unvisited zone by zone and departure hour, etc.).

A key observation supporting the here presented methodology is the fact that the vast majority of the population begins and ends their daily activity schedules at home (Bowman 1998; F. Schneider et al. 2020; Schoenfelder and Axhausen 2001). Consequently, the trips included in ODs should belong to longer home-based trip-chains (often referred as *tours*). This attribute holds particularly true for ODs deriving from urban sensing data sources (e.g. mobile phone data, GPS, etc.) since they are built by continuously tracking the movements of mobile phone holders. This study exploits this observation and attempts to synthesise continuous mobility traces, in the form of tours, based on aggregate ODs.

The identification of continuous mobility traces within ODs is accomplished in a modular fashion as previously suggested by the authors (Ballis and Dimitriou 2020a; 2020c; 2020b). In the case that additional information describing the characteristics of the expected mobility traces is available, additional constraints can be enforced to the optimiser, enabling the identification of a solution which respects the calibration information.

2. Methodology

Overview

This section focuses on describing the suggested methodological framework to create disaggregate mobility traces based on the outputs of typical 4-step model, namely multi-period OD matrices and links counts on the road network. The methodology presented in this paper extends the previously presented framework (Ballis and Dimitriou 2020a; 2020c; 2020b) which enables the conversion of the multi-period ODs to disaggregate home-based, multi-leg, trip-chains (*referred as tours*). For the completeness of the presentation the outline of that framework is also presented here. Firstly, the *graph generation module* exploits the connectivity matrices of the input OD matrices for the conversion of the latter into a single, directed graph. This conversion allows the expression of trip-chains as sequences of zones (i.e. paths) where each edge of the path is directly associated with a trip from the inputted ODs. The second step, referred as the *identification module,* exploits this concept, and applies a sophisticated graph theory-based process for the efficient enumeration of all the possible tours within the graph. The output of this step is a set of all the possible tours that can take place using the trips of the input ODs. For ODs of realistic scale (e.g. hundreds of OD pairs), the enumeration of tours in a graph can prove particularly challenging (mainly due to the combinatorial explosion issue) but the predictability of travel behaviour patterns (C. M. Schneider et al. 2013) can significantly simplify the process. For example, tours with unrealistic durations or with an excessive number of visited zones (e.g. above 6) can be excluded from the search space and as a result support the tackling of combinatorial explosion. The completion of the second step produces a set of plausible tours which are referred as the *candidate tours.* The here presented study takes the approach one step forward by exploiting the traffic counts (i.e. vehicular volumes on each link of the modelled road network) which usually accompany typical 4-step transport models. In particular, the *routing module* is responsible for the translation of the sequences of zones of each candidate tour into likely routes on the road network and subsequently creates the set of *candidate routes.* These candidate routes are provided to the *optimisation module* which deploys a combinatorial optimisation method to identify the combination between the candidates which reproduces the traffic conditions as described in the available traffic counts. More precisely, the optimiser attempts to identify a non-negative frequency for each candidate tours so that the difference between the total number of trips in the inputted ODs and the total number of trips required for the completion of the tours in the solution is minimised.

Mathematical formulation

Let T be the set of candidate tours able to be completed using the trips available in a set of ODs (O). In addition, let R_t be the set of available routes on the (road) network which can be followed to complete a tour t. The optimisation objective (Eq.1) is to identify the frequency of each route ($r \in R_t$) so that the difference between the number of trips required to complete these routes and the available trips in the ODs is minimized. Inequality (Eq.2) guarantees that trips are not used more than their availability, described in the ODs. Moreover, (Eq.3) guarantees that the observed traffic counts are not violated and finally (Eq.4) that the number of times each route is used is non-negative. The optimisation problem can be formulated as:

$$\min Z = \sum_{o \in \mathrm{O}} \left(\sum_{p \in \mathrm{P}} \left(\sum_{t \in T} \left(N_r^t D_t^p \right) \right) - I_p^o \right) \tag{1}$$

Subject to:

$$\sum_{o \in O} \left(\sum_{t \in T} \left(N_r^t D_t^p \right) - T_p^o \right) \leq 0 \;\; \forall \, p \in P \tag{2}$$

$$\sum_{t \in T} \sum_{r \in R_t} \left(N_t^r E_r^l \right) - V_l \leq 0 \; \forall \, \mathrm{l} \in \mathrm{L} \tag{3}$$

$$N_t^r \geq 0 \; \forall \, \mathrm{t} \in \mathrm{T} \tag{4}$$

The abovementioned optimisation formulation is expressed as an integer linear programming problem and is solved here by a Branch and Bound algorithm (IBM 2020). Given a set of multi-period OD matrices and the respective link counts in the modelled road network, the application of the methodology described above, results in a set of disaggregate tours whose included trips represent the input ODs as accurately as possible. This novel tours' synthesis framework is tested on the input provided by a typical 4-step model for the region of Larnaca, Cyprus. The next section presents the details and the performance of the methodology in preparing disaggregate diurnal tours out of aggregate data.

Case study

Overview

The Greater Larnaca Transport Model (GLTM) is a 4-step model developed in VISUM software. The model addresses both private and public transport demand of the morning (AM) and evening (PM) winter weekday peak periods. It has also been designed as a macro simulation Variable Demand Model, thus able to model responses such as changes in the overall transport demand or changes in trip patterns. The study area of GLTM (Figure) has a regional extent and covers the municipality of Larnaca alongside all the nearby suburb areas resulting in a total of 352 zones. The model's degree of precision increases towards the city centres of the urban areas included in this area. All simulations and assignments have been executed on the road network provide by Open Street Maps (OSM) which includes 10,136 nodes and 25,748 links. In terms of travel demand, the model includes multiple modes of transport but for the simplification of the analysis, the focus has been placed only on the private car trips. The total number of car trips in the AM period (07:00-09:00) are estimated at 31,132 while for the PM peak (04:00-06:00) at 17,592 indicating that the morning peak is considerably more prominent (Figure). In addition, the difference in trips between the AM and PM periods indicates that the return trips for tours beginning in the AM period do not take place during the PM period, therefore the ability of the suggested method to produce diurnal tours is highly constrained.

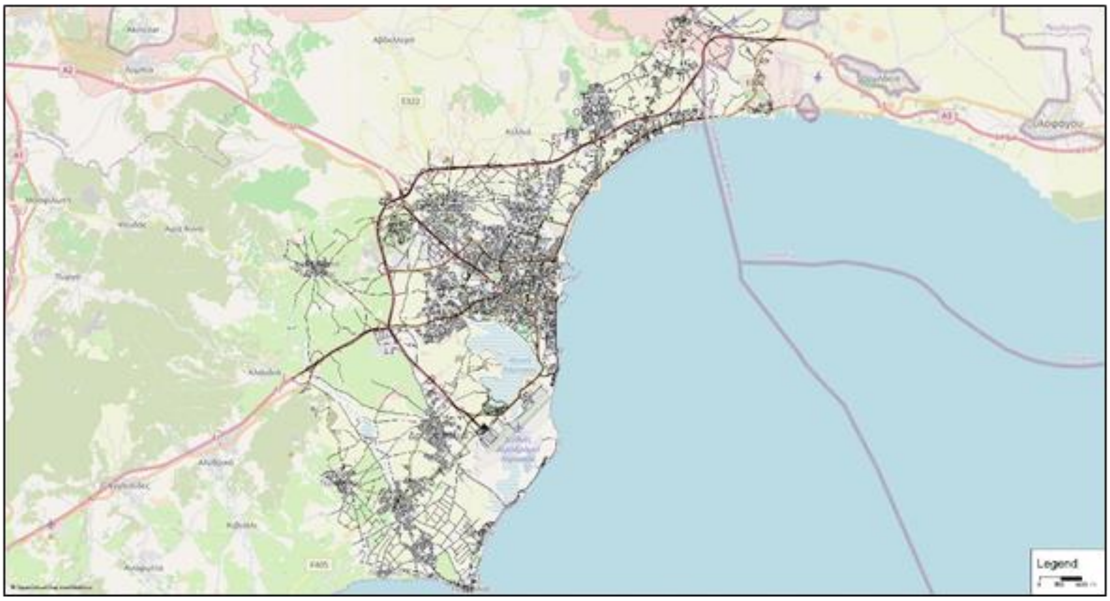

Figure 1. Study Area of GLTM

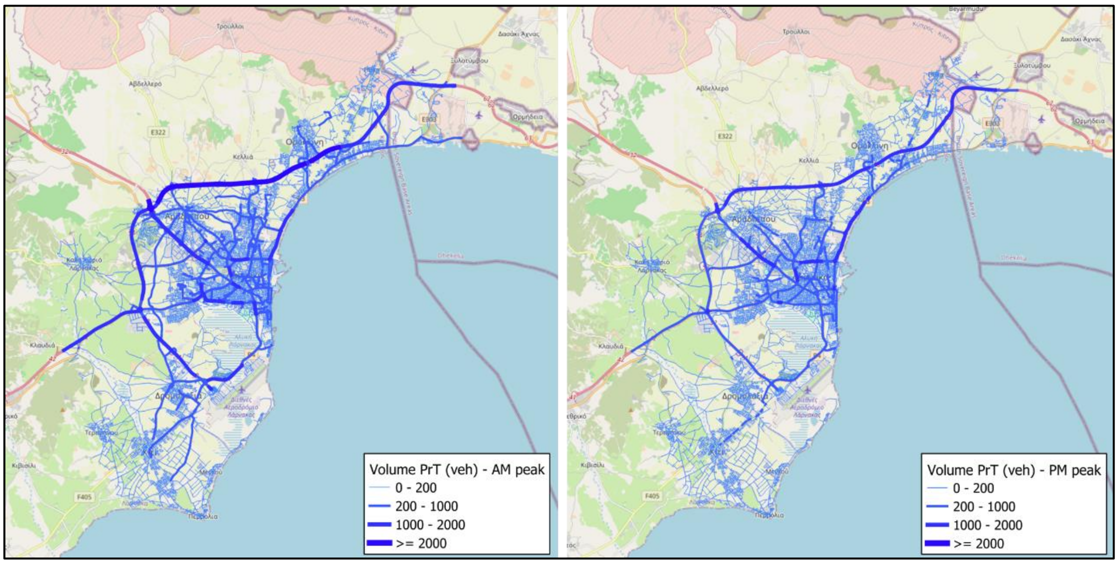

Figure 2. Link volumes for the AM (left) and PM (right) peak periods

Results

The next section delves in the presentation of the results obtained from the application of the previously described methodology on the 4-step model of Larnaca.

Validation

This section presents the results concerning the accuracy of the approach. Although the methodology has not managed to fully convert the available OD trips into multi-leg tours, the results can be considered as accurate and more importantly enable the significant increase the value of the original, aggregate ODs. In detail, the process has managed to include 21,808 out of the 48,724 trips (44%) of the observed ODs into 5,088 unique tours using 7,762 different routes for their completion. This relatively low percentage can be explained to a large extent by the non-symmetric nature of the observed ODs. As shown in Figure 1, the number of trips originating and ending at the zones of the observed ODs is uneven for the majority of zones, therefore the ability of the method to combine trips into complete tours is significantly hindered. Given the fact that the methodology did not overestimate the existence of tours within the inputted ODs, it can be argued that the approach is non-susceptible to inconsistent input and in addition that it can be used as an effective mechanism to validate the existence of tours in ODs.

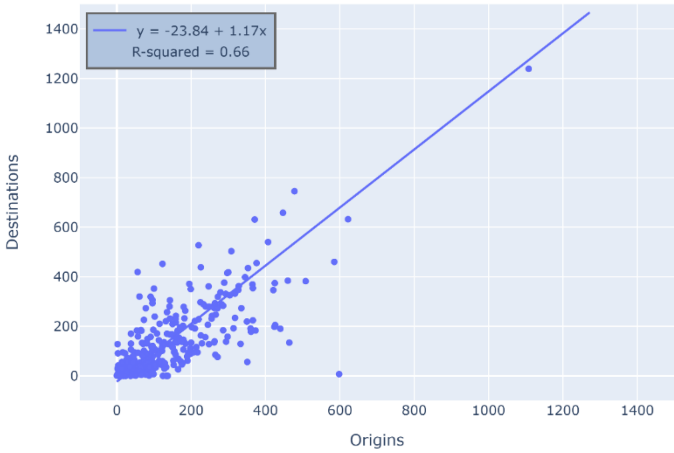

Figure 1. Evaluation of the symmetry between trip origins and trip destinations for the observed ODs

The relatively low number of tours included in the modelled solution (i.e. modelled ODs) is also reflected in the comparison against the observed OD with Figure 2 presenting the correlation between them. As it can be noted, the correlation between the trips in the cells of the observed and the modelled ODs although low, it is still considerable, and a significant number of trips has been incorporated into multi-leg tours.

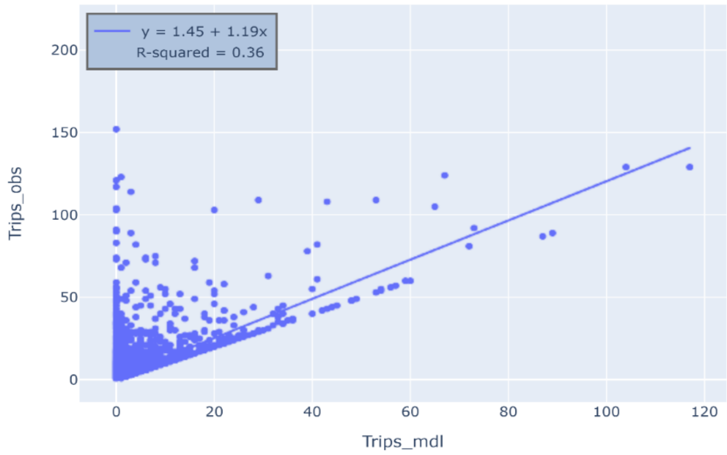

Figure 2. Scatter matrix comparison for the trips of the observed and the modelled OD matrices

Even though the total number of trips in the modelled ODs is not equivalent to the total number in the observed ODs, the methodology has managed to transform a significant portion of the inputted ODs into likely tours. The likeliness of the modelled tours and the accuracy of the model is reinforced through the exploitation of the available link traffic counts (provided by the 4-step model of Larnaca). By imposing the traffic counts as constraints, the methodology achieves the identification of tours and their routes which confront to the observed traffic patterns. Figure 3 presents the observed traffic counts versus the counts deriving from the aggregation of the modelled routes. Although the total number or link counts differs significantly between the observed counts (2,799,013 vehicles) and modelled counts (992,682 vehicles), the results present high correlation (R^2=0.72) and indicate that the modelled routes follow the observed traffic patterns to a high degree. More importantly, the results indicate that the modelled tours are translated to realistic routes on the network, therefore the tours reproduce realistic patterns of traffic.

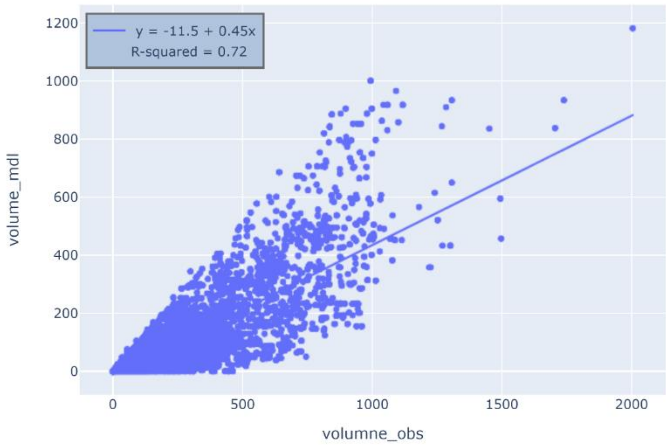

Figure 3. Scatter matrix comparing the observed versus the modelled link counts

3. Conclusions

The presented study reinforced the ability of the previously presented methodological framework (Ballis and Dimitriou 2020a; 2020c; 2020b) to convert aggregate ODs into disaggregate and fully tractable home-based trip-chains (i.e. tours). One main requirement of this methodology is the provision of calibration information which enable the identification of the most likely solution out of all the possible ones. This study takes the framework one step forward by exploiting observed traffic counts for the provision of the required calibration data. In particular, the study uses aggregate ODs and traffic counts deriving from a 4-step developed for the city of Larnaca, Cyprus and converts them into realistic multi-leg, tours. The methodology managed to convert almost 45% of the trips included in the observed ODs into such tours which closely follow the observed traffic patterns. Based on the non-symmetry characterising the observed ODs (i.e. the number of originating trips for many zones does not match the number of trips ending to them), it can be argued that the methodology has correctly excluded the rest of the observed trips (55%) instead of incorporating them into unrealistic tours.

REFERENCES

Anda, Cuauhtemoc, Sergio Ordonez Medina, Sergio Arturo, Kay W Axhausen, and Sergio A Ordonez Medina. 2020. “Synthesising Digital Twin Travellers Individual Travel Demand from Aggregated Mobile Phone Data Synthesising Digital Twin Travellers: Individual Travel Demand from Aggregated Mobile Phone Data.” Zurich, Switzerland. https://doi.org/10.3929/ethz-b-000442517.

Ballis, Haris, and Loukas Dimitriou. 2020a. “Deriving Daily Activity Schedules from Dynamic, Purpose Dependent Origin-Destination Matrices.” *Transportation Research Record* 2674 (10): 375–86. https://doi.org/10.1177/0361198120939094.

———. 2020b. “Optimal Synthesis of Tours from Multi-Period Origin-Destination Matrices

Using Elements from Graph Theory and Integer Programming.” *European Journal of Transport and Infrastructure Research* 20 (4): 1–21. https://doi.org/10.18757/ejtir.2020.20.4.5303.

———. 2020c. “Revealing Personal Activities Schedules from Synthesizing Multi-Period Origin-Destination Matrices.” *Transportation Research Part B: Methodological* 139 (September): 224–58. https://doi.org/10.1016/j.trb.2020.06.007.

Ben-Akiva, Moshe, Jon Bottom, Song Gao, Haris N. Koutsopoulos, and Yang Wen. 2007. “Towards Disaggregate Dynamic Travel Forecasting Models.” *Tsinghua Science and Technology* 12 (2): 115–30. https://doi.org/10.1016/S1007-0214(07)70019-6.

Bhat, Chandra R., and Frank S. Koppelman. 1999. “Activity-Based Modeling of Travel Demand,” no. January 2003: 35–61. https://doi.org/10.1007/978-1-4615-5203-1_3.

Goulias, Konstadinos G. 2009. “Travel Behavior and Demand Analysis and Prediction.” In *Encyclopedia of Complexity and Systems Science*, 9536–65. Springer New York. https://doi.org/10.1007/978-0-387-30440-3_565.

Hilgert, Tim, Martin Kagerbauer, Thomas Schuster, and Christoph Becker. 2016. “Optimization of Individual Travel Behavior through Customized Mobility Services and Their Effects on Travel Demand and Transportation Systems.” *Transportation Research Procedia* 19 (June): 58–69. https://doi.org/10.1016/j.trpro.2016.12.068.

Horn, M.E.T. 2002. “Multi-Modal and Demand-Responsive Passenger Transport Systems: A Modelling Framework with Embedded Control Systems.” *Transportation Research Part A: Policy and Practice* 36 (2): 167–88. https://doi.org/10.1016/S0965-8564(00)00043-4.

Huber, Stefan, and Sven Lißner. 2019. “Disaggregation of Aggregate GPS-Based Cycling Data – How to Enrich Commercial Cycling Data Sets for Detailed Cycling Behaviour Analysis.” *Transportation Research Interdisciplinary Perspectives* 2 (December): 100041. https://doi.org/10.1016/j.trip.2019.100041.

Kamargianni, Maria, Melinda Matyas, Weibo Li, and Andreas Schäfer. 2015. “Feasibility Study for ‘ Mobility as a Service ’ Concept in London,” 84. https://doi.org/10.13140/RG.2.1.3808.1124.