VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Vytautas DUMBLIAUSKAS

DEVELOPMENT AND APPLICATION OF TOUR-BASED TRAVEL DEMAND MODEL FOR PLANNING OF URBAN TRANSPORT NETWORKS

DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,
TRANSPORT ENGINEERING (T 003)

VGTU
LEIDYKLA
Vilnius TECHNIKA 2019

This doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2014–2019.

Supervisor

Assoc. Prof. Dr Vytautas GRIGONIS (Vilnius Gediminas Technical University, Transport Engineering – T 003).

The Dissertation Defence Council of Scientific Field of Transport Engineering of Vilnius Gediminas Technical University:

Chairman

Prof. Dr Gintautas BUREIKA (Vilnius Gediminas Technical University, Transport Engineering – T 003).

Members:

Dr Dago ANTOV (Tallinn University of Technology, Estonia, Transport Engineering – T 003),
Prof. Dr Romas BARONAS (Vilnius University, Informatics – N 009),
Assoc. Prof. Dr Artūras KERŠYS (Kaunas University of Technology, Transport Engineering – T 003),
Prof. Dr Audrius VAITKUS (Vilnius Gediminas Technical University, Civil Engineering – T 002).

The dissertation will be defended at the public meeting of the Dissertation Defence Council of Transport Engineering in the Senate Hall of Vilnius Gediminas Technical University at **9 a. m. on 12 April 2019**.

Address: Saulėtekio al. 11, LT-10223 Vilnius, Lithuania.
Tel.: +370 5 274 4956; fax +370 5 270 0112; e-mail: doktor@vgtu.lt

A notification on the intended defending of the dissertation was sent on 8 March 2019.

A copy of the doctoral dissertation is available for review at VGTU repository http://dspace.vgtu.lt, at the Library of Vilnius Gediminas Technical University (Saulėtekio al. 14, LT-10223 Vilnius, Lithuania), at the library of Klaipėda University (K. Donelaičio st. 3, LT-92144 Klaipėda, Lithuania), at the library of Vytautas Magnus University (K. Donelaičio st. 52-215 rm., LT-44244 Kaunas, Lithuania) and at the library of Kaunas University of Technology (K. Donelaičio st. 20, LT-44239 Kaunas, Lithuania).

VGTU leidyklos TECHNIKA 2019-004-M mokslo literatūros knyga

ISBN 978-609-476-156-0

vytautas.dumbliauskas@vgtu.lt

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Vytautas DUMBLIAUSKAS

KELIONIŲ GRANDINĖMIS PAGRĮSTO SUSISIEKIMO POREIKIŲ MODELIO KŪRIMAS IR TAIKYMAS MIESTŲ SUSISIEKIMO TINKLO PLANAVIMUI

DAKTARO DISERTACIJA

TECHNOLOGIJOS MOKSLAI,
TRANSPORTO INŽINERIJA (T 003)

VGTU LEIDYKLA

Vilnius TECHNIKA 2019

Disertacija rengta 2014–2019 metais Vilniaus Gedimino technikos universitete.

Vadovas

doc. dr. Vytautas GRIGONIS (Vilniaus Gedimino technikos universitetas, transporto inžinerija – T 003).

Vilniaus Gedimino technikos universiteto Transporto inžinerijos mokslo krypties disertacijos gynimo taryba:

Pirmininkas

prof. dr. Gintautas BUREIKA (Vilniaus Gedimino technikos universitetas, transporto inžinerija – T 003).

Nariai:

dr. Dago ANTOV (Talino technologijos universitetas, Estija, transporto inžinerija – T 003),
prof. dr. Romas BARONAS (Vilniaus universitetas, informatika – N 009),
doc. dr. Artūras KERŠYS (Kauno technologijos universitetas, transporto inžinerija – T 003),
prof. dr. Audrius VAITKUS (Vilniaus Gedimino technikos universitetas, statybos inžinerija – T 002).

Disertacija bus ginama viešame Transporto inžinerijos mokslo krypties disertacijos gynimo tarybos posėdyje **2019 m. balandžio 12 d. 9 val.** Vilniaus Gedimino technikos universiteto senato posėdžių salėje.

Adresas: Saulėtekio al. 11, LT-10223 Vilnius, Lietuva.
Tel.: (8 5) 274 4956; faksas (8 5) 270 0112; el. paštas doktor@vgtu.lt

Pranešimai apie numatomą ginti disertaciją išsiusti 2019 m. kovo 8 d.

Disertaciją galima peržiūrėti VGTU talpykloje http://dspace.vgtu.lt ir Vilniaus Gedimino technikos universiteto (Saulėtekio al. 14, LT-10223 Vilnius, Lietuva), Klaipėdos universiteto (K. Donelaičio g. 3, LT-92144 Klaipėda, Lietuva), Vytauto Didžiojo universiteto (K. Donelaičio g. 52-215 kab., LT-44244 Kaunas, Lietuva), Kauno technologijos universiteto (K. Donelaičio g. 20, LT-44239 Kaunas, Lietuva) bibliotekose.

Abstract

This thesis is devoted to the analysis of the advanced and innovative tour-based travel demand modelling approach. Tour-based models explicitly recognise traffic as a derived demand for undertaking activities between homes and destinations. Travel demand of urban residents is modelled as trip sequences, which allows precise modelling of trip origin and destination points. The tour-based approach is deemed as a key step forwards towards even more complex agent-based modelling systems.

The thesis is structured around three main chapters that can be summarised succinctly as a revision of the state of the practice and research, description of empirical research of travel behaviour, and tour-based model development.

The 1st chapter revises the current state of practice and the research on travel demand modelling. All the building blocks that comprise transport models are discussed, and this lays the theoretical foundation for the following chapters. 1st chapter also gives a thorough comparison of trip-based and tour-based modelling approaches and presents modelling environment.

The 2nd chapter defines the process of conducting an empirical research of the travel behaviour patterns of urban residents. The 2nd chapter defines survey methodology and important mobility parameters such as activity sequences and their probabilities of homogeneous urban population segments. The outputs from the 2nd chapter are not only important and interesting on their own, but they also flow into the final third part of the work.

The final 3rd chapter defines tour-based travel demand model development steps and showcases their practical application to the real-world scenario. Demand model quality assessment efforts and results are presented and discussed together with necessary explanations for significant deviations from reality. The resulting model is applied to investigate the performance of Siaurine Street in Vilnius, which is to be built in the coming years. At the very end of 3rd chapter a comprehensive urban travel demand modelling framework is formulated and serves as a best practice guide.

General conclusions summarises the whole study. These are followed by an extensive list of references that were mentioned or relied upon to some extent in the work. Finally, separate lists of scientific publications and conference presentations conclude the thesis.

Overall, there have been five scientific articles published on the topic of the thesis. Four articles were published in scientific journals that are referenced in Clarivate Analytics Web of Science database, and one article was published in a scientific journal that is referenced in other databases.

Reziumė

Disertacijoje nagrinėjami kelionių grandinėmis pagrįsto susisiekimo poreikių modelio kūrimas ir taikymas. Kelionių grandinėmis pagrįstame modelyje keliavimas apibrėžiamas kaip poreikis, kylantis iš noro atlikti skirtingas veiklas skirtingose geografinėse miesto vietose. Miesto gyventojo kelionės modelyje aprašomos kaip kelionių grandinės, kurių pagalba parenkama tiksli modeliuojamų kelionių pradžia ir pabaiga. Kelionių grandinėmis pagrįsta modeliavimo metodika yra tarpinis žingsnis link dar pažangesnių ir sudėtingesnių veiklomis pagrįstų modelių.

Disertaciją sudaro trys pagrindiniai skyriai, kurie apibendrintai gali būti pavadinti taip: literatūros apžvalga, empiriniais duomenimis pagrįsta susisiekimo poreikių analizė, kelionių grandinėmis pagrįsto modelio kūrimas ir šio modelio eksperimentinio taikymo aprašymas.

Pirmajame disertacijos skyriuje apžvelgiami teoriniai ir praktiniai susisiekimo poreikių modelių aspektai. Šiame skyriuje taip pat glaustai palyginama įprastinė modeliavimo paradigma su kelionių grandinėmis pagrįsta paradigma. Modeliavimo aspektai aptarti pirmame skyriuje formuoja teorinį pagrindą tolimesniems skyriams.

Antrajame disertacijos skyriuje yra teikiama miesto gyventojų keliavimo įpročių empirinės analizės metodika. Skyriuje aprašoma anketinės apklausos kūrimo metodika ir analizuojami svarbūs gyventojų judumo parametrai, tokie kaip homogeniškų miesto gyventojų grupių atliekamos veiklų sekos ir jų tikimybės. Šiame skyriuje atlikto tyrimo rezultatai yra svarbi kelionių grandinėmis pagrįsto susisiekimo poreikių modelio dalis.

Trečiajame skyriuje yra detalizuojami modelio kūrimo žingsniai, teikiami inovatyvūs pasiūlymai dėl šių žingsnių tobulinimo ir demonstruojamos praktinio pritaikymo galimybės. Skyriuje pristatoma sukurta susisiekimo poreikių modelio reprezentatyvumo užtikrinimo metodika ir įvertinamas sukurto modelio patikimumas. Skyriaus pabaigoje teikiama apibendrinanti susisiekimo poreikių rengimo metodika, kuri gali būti gerosios praktikos vadovu.

Pabaigoje yra pateikiamos bendrosios išvados, kurios apibendrina visus darbo rezultatus. Po išvadų pateikiamas cituotos literatūros sąrašas ir autoriaus mokslinių publikacijų disertacijos tema sąrašas.

Disertacijoje apibendrinti tyrimai buvo paskelbti penkiuose straipsniuose. Keturi straipsniai buvo paskelbti recenzuojamuose mokslo žurnaluose, kurie yra reitinguojami Clarivate Analytics Web of Science duomenų bazėje, o vienas straipsnis buvo publikuotas mokslo žurnale, kuris yra reitinguojamas kitose duomenų bazėse.

Notations

Main Definitions

Activity – an occupation of a person undertaken at one location.

Activity Sequence – an ordered list of activities undertaken over a person´s typical day, starting and ending at home.

Agent-based Travel Demand Model – a model that simulates individual agents making travel decision associated with a person’s activities over the course of the day.

Stage – a continuous movement with one mode of transport or one vehicle.

Tour – a sequence of trips comprising the home-to-home loop.

Tour-based Travel Demand Model – a model comprised of three separate steps (trip generation, trip distribution and mode choice, assignment) treating person’s trips as interdependent units of travel.

Travel Demand Model – a computational procedure that estimates travel demand.

Trip – a continuous sequence of stages between two activities.

Trip-based Travel Demand Model – a model comprised of four separate steps (trip generation, trip distribution, mode choice, assignment) treating a person’s trips as independent and unrelated units of travel.

Abbreviations

ABM – Agent-based Modelling;

AD – Advantage Distance;

AM – Before Midday;

ANT – Average Number of Trips;

API – Application Programming Interface;
ARR – Arrival Time of Connection;
AST – Activity Start Time;
AT – Access Time;
ATP – Average Trip Length;
ATR – Auxiliary Transport Ride Time;
BD – Bachelor Degree;
CR – Correction Factor;
CS – Connection Segment;
D – Distance-based Modal Split;
DEP – Departure Time of Connection;
DT – Daily Tours;
DTP – Distribution of Trip Purposes;
E – Employed;
ET – Egress Time;
F – Female;
FSM – Four-step Modelling;
GEH – Statistic invented by Geoffrey E. Havers;
GIS – Geographic Information System;
GTFS – General Transit Feed Specification;
HE – Higher Education;
IMP – Impedance;
IVT – In-Vehicle Time;
JT – Journey Time;
LLC – Limited Liability Company;
M – Male;
MS – Mode Share;
NDS – Number of Daily Sequences;
NPC – Network Performance Characteristics;
NT – Number of Transfers;
OD – Origin and Destination;
OSM – Open Street Map;
PhD – Doctor of Philosophy;
PJT – Perceived Journey Time;
PF – Passenger Flows;
PM – Past Midday;

PT – Proportion of Travelling Respondents;
PSE – Primary or Secondary Education;
R – Retired;
RS – Route Segment;
S – Student;
T – Trip-based Modal Split;
TAZ – Transport Analysis Zone;
TDM – Travel Demand Management;
TT – Travel Time;
TWT – Transfer Wait Time;
U – Unemployed;
VE – Vocational Education;
VF – Vehicle Flows;
VGI – Volunteered Geographic Information;
WebTAG – Web-based Transport Analysis Guidance;
WGS84 – World Geodetic System Established in 1984;
WKT – Walk Time.

Contents

[1] The annexes are provided in the enclosed disc.

Introduction

Problem Formulation

Transport planning methods in Lithuania have been advancing quite rapidly since the declaration of independence in 1990. Until 2015 the main methodological guide for planning and design was "Transport Systems of Cities, Towns and Villages", which was then replaced by The Lithuanian Ministry of Environment into two separate guidelines. Guidelines entitled "Streets and Local Roads" now detail the design procedures, whereas "Urban Transport Planning Guidelines" are generally applicable to spatial planning procedures. In addition, with incentives and support from European Commision, the Lithuanian Ministry of Transport and Communications has started focusing on sustainable urban mobility. To this end, it has approved Sustainable Urban Mobility Guidelines, which describe the required content of Sustainable Urban Mobility Plans, the planning procedure and the roles of various stakeholders. Inevitably, new set of legally binding documents will have a significant and hopefully positive impact on the planning practice and the transport system itself.

However, the new set of regulatory documents do not ensure satisfactory basis for analytical decision making processes. "Urban Transport Planning Guidelines" recognize the need to apply transport modelling techniques whilst developing Master Plans of towns and cities with more than fifty thousand residents. On the other hand, a sole statement about the need to use such a methodology does not garantuee desired outcomes, because there are no guidelines detailing the exact and widely accepted methodology.

Relevance of the Thesis

Due to foregoing, prioritisation of investments for major network developments currently is based solely on the political and engineering intuition most of the time. To ensure a fair fulfilment of society’s needs, and the rational use of available resources, this process must be accompanied by independent analytical decision-making techniques, such as travel demand modelling paired with cost-benefit analysis. The United Kingdom, for instance, has a rather sophisticated set of guidelines, which have far reaching traditions and are widely applied in the field.

This thesis embarks on an endeavour to create a transport modelling methodological framework, which would be based on the current best practices and state of the research, and which at the same time, could be accessible by a wide range of transport specialists.

The Object of the Research

The research object is urban transport system’s components: travel demand and travel supply.

The Aim of the Thesis

To develop an innovative urban transport modelling methodological framework ensuring representation of travel supply and demand for the whole day.

The Objectives of the Thesis

To achieve the aim of the thesis, the following objectives were formulated:

1. To compare and contrast urban transport modelling methodologies prevalent in theory and practice.
2. To develop and test efficient and up-to-date data collection techniques ensuring representation of travel supply and travel demand for the whole day.

3. To create a travel demand model representing daily demand and supply for planning of urban transport system and apply it in the assessment of the impact of the transport network development scenarios.
4. To develop a comprehensive methodological framework for urban transport modelling, which would guarantee statistically reliable models and could be easily applicable for planning of transport networks.

Research Methodology

Along the path towards this final thesis aim, a few widely recognized methods and tools were employed:

1. Systematic review of scientific literature to summarise strengths and weaknesses of modelling methodologies.
2. Empirical travel behaviour survey and secondary travel distance data collection by means of Google Distance Matrix API were employed in gathering information about respondents’ mobility patterns.
3. Python programming language and its standard libraries such as Pandas, NumPy, SciPy, Seaborn etc. were used to undertake mining, cleaning and statistical analysis of data further utilized for the representation of demand and supply within travel demand model.
4. An open source geographic information system QGIS was employed to accomplish geospatial analysis tasks and facilitate development of travel supply and demand datasets for the representation within travel demand model.
5. PTV Visum modelling environment was employed for transport supply and travel demand data management, analysis and modelling.

Scientific Novelty of the Thesis

The novelty of this thesis is characterised by the following aspects:

1. Innovative empirical travel behaviour data collection, processing and analysis approach allowing the identification of activity sequences and their probabilities by homogeneuous population segments. This novel approach can be harnessed in the scientific and applied travel behaviour studies.
2. The new and open source datasets such as Open Street Map (OSM) and General Transit Feed Specification (GTFS) was applied within the travel

supply model building process. This ensures a precise representation of the transport networks and reduces model development cost.

3. Development and application of the innovative tour-based travel demand model with a help of up-to-date data mining and processing technologies.
4. The formulation of a comprehensive urban transport modelling methodological framework, which, if applied by practitioniers, ensures a consistent and robust transport modelling approach.

Practical Value of the Research Findings

An innovative data processing techniques were applied for the development of travel demand and transport supply within tour-based model. Tour-based model allows evaluation of various scenarios related to the development of a transport system (new public transport modes, park and ride system, etc.) or the travel management (new urbanized areas, spatial distribution of activity centres, etc.). In addition, the tour-based model was applied to assess the impact of Siaurine Street to travel demand and network performance.

On the other hand, a comprehensive urban transport modelling framework is universal and transferable to the context of any other urban territory without restrictions to the national context. However, additional data collection, manipulation, analysis and modelling efforts will be necessary.

The Defended Statements

1. Tour-based travel demand model is superior to trip-based travel demand model, because of the coverage of a larger set of trip purposes, a capability to take into account daily trip sequences and a consistent mode choice estimation within the tour.
2. The national institutionalization of consistent and recurring travel behaviour studies, employing an activity sequence-based approach, coupled with advanced data collection technologies, would create a basis for retrospection, would ensure a detailed identification of travel behaviour patterns across homogeneous population segments and would guarantee the capability of representing daily demand within travel demand models.
3. The application of open source datasets within the travel supply and demand model development ensures detailed public and private transport network representation and financial cost reduction.

4. A comprehensive urban transport modelling framework is essential for a statistically reliable and consistent development of transport models. Hence, the proposed architecture can serve as the best practice guide ensuring desired modelling characteristics.

Approval of the Research Findings

Five articles in total have been published on the research topic in the scientific journals. Four articles were published in periodic scientific journals that are referenced in Clarivate Analytics Web of Science database. One article was published in a periodic scientific journal that is referenced in other databases.

Research results were presented at two international scientific conferences:

1. Barauskas, A; Dumbliauskas, V. 2016. A Study on the Travel Time Generalized Cost Function in Kaunas City. 28th European Conference of Operational Research (Poznan University of Technology).
2. Dumbliauskas, V; Grigonis, V. 2017. Estimating the Effects of Public Transport Priority Measures at Signal Controlled Intersections. 10th International Conference of Environmental Engineering (VGTU).

Structure of the Dissertation

The dissertation is structured around three main chapters that can be summarised succinctly as: a review of travel demand modelling techniques; travel demand empirical research (methodology, data collection and statistical analysis); and tour-based model development.

The overall work fits in 126 pages and consists of 90 references, 32 figures, 21 tables and 43 numbered formulas.

1

A Review of Travel Demand Model Development Techniques

Understanding and solving transport problems is a process which generally starts with an analysis of the current state and the purpose of identifying deficiencies. Further, the analysis of the current state is followed by a design process with the purpose of removing deficiencies and potentially improving the system significantly.

Transport modelling is a procedure that supports the current state analysis as well as generating the evaluation of various development scenarios. Very frequently, the results of transport modelling are a key input to the decision making at the political level.

In Lithuania, unfortunately, modelling is not applied frequently because of two main reasons:

1. There is no proper methodological and legal background. In other words, legislation has not made modelling mandatory and has not created any methodological guidelines.
2. There is a lack of expertise in the field. As the demand for studies involving travel demand modelling is very low, the expertise has not evolved over time and in many cases, has moved to other markets.

As a result, the development of transport systems is currently missing some potential advantages that can be brought in by a comprehensive urban transport modelling process. These are:

1. Political decisions regarding transport investments would be supported by scientifically valid methodology and would have more transparent and robust justification.
2. The analysis would provide quantitative performance metrics in transport related (traffic flows, travel times etc.) and/or economic related (cost-benefit ratio, net present value, internal rate of return) terms.

On the other hand, it is worthwhile saying that modelling involves some apparent drawbacks. For instance, it frequently entails a number of uncertain assumptions, which reduces credibility within the modelling outcomes.

Transport models consist of two fundamental submodels: travel demand and network supply. Both have very different underlying principles as travel demand is closely related to the population behaviour whereas the network supply model reflects the physical transport infrastructure and operation.

Two scientific articles (Grigonis *et al.* 2014; Dumbliauskas *et al.* 2018) have been published on the topic of this chapter.

1.1. Travel Demand Modelling Techniques

1.1.1. General Review

Travel demand is a secondary demand, arising from the spatial separation of home and basic human social activities such as work, education, shopping and recreation. Individuals travel from one point to another with trips of different purposes by various modes and durations and at various times of the day.

The diversity of individual human behaviour across geographical locations ensures that traffic behavioural data is not universal. Behavioural characteristics vary widely between different cities (and even different areas within the same city) and while there are some general similarities that can be found in comparisons of cities, there are many different factors that influence human travel behaviour, including: the size of the city, its urban density, its layout, the demographic and cultural properties of its population, economic conditions and the type and quality of the transport networks. All these factors play a vital role in influencing transport demand. Travel demand modelling techniques allows quantification and further analysis of travel demand by taking into account various of the above mentioned factors.

In summary, there are three fundamentally different approaches towards modelling of large scale urban transport systems and they can be listed in the following chronological order:

1. Trip-based travel demand models.
2. Tour-based travel demand models.
3. Agent-based travel demand models.

Trip-based travel demand modelling procedure (a.k.a conventional travel demand modelling procedure or four step approach) that was first developed in 1950s has become a de facto standard among the practitioners. But since its introduction, some significant problems have been recognized by scientists and practitioners and have become common knowledge:

1. The need to aggregate trip origins and destinations into analysis zones introducing aggregation errors and the related problem of dealing with intrazonal trips.
2. The treatment of individual trips as independent decisions where the effects of other activity decisions are not considered (even the mode choice for a preceding trip to the origin of current trip).
3. A difficulty in analysing future travel behaviour that is rare or does not exist at present, including for example behaviour that might result from the implementation of new policies such as congestion pricing.
4. A difficulty in modelling time of day related issues, especially departure time choice and peak spreading.

The inability of the trip-based procedure to perform certain necessary transportation analyses required today when the emphasis is on travel demand management (a.k.a mobility management) together with recent advances in both research and computing capability has prompted a re-examination of the paradigm (Rossi *et al.* 1997).

A step towards more sophisticated systems has been taken with the advent of tour-based travel demand model. Tour-based travel demand model, where the unit of travel is defined as the tour from home to one or more destinations and then back home, is a reasonable near-term alternative to the conventional trip-based process, where the individual trip is the unit. Tour-based travel demand model carries an assumption that travel demand is derived from the desire for activities at physically separate locations.

While all the issues listed above cannot be addressed by a tour-based model system, it can address the second issue listed above, which is one of the most critical ones. As a practical alternative, the tour-based demand model deals with the sequences of activities and tours and consider the interlocking dependencies between trips. However, the tour-based models are only a temporary solution towards more sophisticated approaches. It is worth recognising that tour-based modelling is a step towards agent-based models that simulate individual agents making travel decision associated with the person's activities over the course of the day.

First, a synthetic population based on census and survey data is built; second, activities and associated locations for all individuals are generated; then, multimodal trips satisfying those activities are estimated by selecting optimal routes; finally, a microsimulation over the entire transportation system of all agents (including transit) is undertaken. The goal is to gain detailed traffic data in a given study area to support traffic, travel demand, and transportation policy analyses (Zheng *et al.* 2013).

Unfortunately, such a detailed representation of the system comes with its own price. The software systems employed to represent the reality are complicated and more often than not require software development skills that are rare among transport planners and modellers. On the other hand, large datasets that are necessary as an input are also not yet widely available and this in turn reduces the opportunities for applications.

Despite these barriers, scientists try to promote this paradigm via open source platforms that are freely available. Good examples are:

1. TRANSIMS (Transportation Analysis and Simulation System) was initially developed at Los Alamos National Laboratory by Nagel *et al.* (1999) with an attempt to build a tool that is completely microscopic, which means that it keeps track of individual travellers and represents transportation infrastructure, such as intersections, traffic lights, turn pockets microscopically. A review of its application cases is given by Lee *et al.* (2014).
2. MATSIM (Multi-Agent Transport Simulation) was started by the same prof. Kai Nagel at ETH Zurich Swiss Federal Institute of Technology in Zurich. Horni *et al.* (2016) provides a complete guide for use of this tool with transportation planning models.

Zhong *et al.* (2015) acknowledge that despite the advances of agent-based modelling in the academic context, conventional trip-based travel demand modelling technique remains the most popular modelling approach and is still being used by the majority of the Metropolitan Planning Organizations in the USA. As it is stated in Vovsha *et al.* (2005), trip-based modelling approach has been established through decades of application and experience and this cannot be simply ignored.

On the other hand, it is important to consider improving the conventional trip-based modelling approach and to take at least some of the advantages, if not all, of the new generation models in order to get effective results and to address problems that cannot be solved using conventional methods (Omer *et al.* 2010).

Therefore, the tour-based travel demand model, having almost the same input data requirement, development efforts and running times as trip-based models as well as offering practical improvements, should be applied in the field in our national context until the agent-based paradigm has achieved its mature stage.

After the review of key definitions, the chapters that follow delve into a more detailed investigation of both trip-based and tour-based approaches.

1.1.2. Key Travel Demand Definitions

Axhausen (2007) points out the need for clear definitions to make sense of the scientific observations and outcomes of survey-based research and transport modelling. The author of this thesis cannot agree more and, therefore, clear main definitions, which will be valid throughout this work, are established within this section.

An activity is defined as an occupation of a person carried out at one location. It is worth differentiating between human needs related activities such as work, shopping or social communication and travel related activities such as change of mode or a transfer between the vehicles of the same mode, which are referred to as a process. A sequence of activities describes the order of different activities during a person´s run of the day, starting and ending at home, for instance, the very frequent sequence undertaken by population members is “Home–Work–Shopping–Home”.

A stage is a continuous movement with one mode of transport or one vehicle. A trip is a continuous sequence of stages between two activities. For example, a trip with public transport usually is defined by at least three separate stages: walk, travel by public transport and walk again.

A tour concept is a key term within the scope of this work. According to Krizek (2003), tours in literature are defined in terms of the home-to-home loop and analysed by looking at the number of trips. Simple tours contain two trips; complex tours contain more than two trips. These terms are employed in the further sections and used extensively.

Other authors (Strathman *et al.* 1994; Wallace *et al.* 2000) tend to name the same concept as a trip chaining. However, in the context of this thesis, a similar term (trip chain) is assigned to a slightly different meaning (see below) and therefore care should be taken to avoid associating the term trip chaining with the meaning of a tour.

These latter three definitions (stage, trip and tour) align with the ones agreed among the bulk of transport planning professionals and given by Ortuzar *et al.* (2011).

A trip chain is a sequence of two or more trips between two substantial activities (i.e home and work). An activity is treated as substantial if it takes place longer than some predefined arbitrary time. Further work will be following Wallace *et al.* (2000), who assumed that activity is substantial if it takes place for longer than 90 minutes. Sometimes a “trip chain” by other authors (Hensher *et al.* 2000; Lee *et al.* 2002) is characterised as travel that almost always begins and ends at home, thus being assigned a meaning of a “tour”.

To facilitate the apprehension, a schematic representation depicting definitions is given in Figure 1.1.

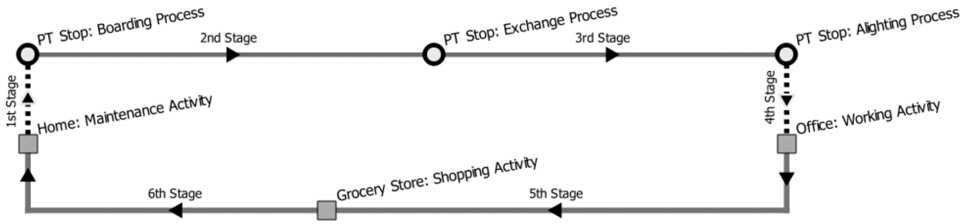

Fig. 1.1. Visualisation of the Key Definitions (Created by Author)

In the diagram presented above, there are six stages that comprise three trips. The first trip consists of four stages between two activities maintenance at home and work at the office. The second trip covers only one stage between work and shopping activities. And finally, the third trip is also defined by one (6^{th}) stage and connects shopping and maintenance (at home) activities. If the shopping is not taking place for longer than 90 minutes we identify two trip chains: first from home to work and second from work to home. Otherwise there exists only one trip chain – between home and work. All six stages form three trips and two trip chains comprise a single tour that starts and ends at home. An ordered list of activities (Maintenance, Working, Shopping, Maintenance) constitute the activity sequence.

1.2. Trip-based Travel Demand Modelling Approach

1.2.1. Main Idea

Four steps of the conventional trip-based transport demand model are:

1. Generation of trip ends, which determines the number of trips for each purpose arising at various origin locations and ending at various destination locations in each zone, calculated as a function of land use, demographics and other socio-economic factors.

2. Trip distribution, which matches trips associated with each origin to each destination, often using a gravity model function.
3. Mode choice computes the proportion of trips between each origin and destination that uses various transportation modes.
4. Assignment, where the estimated demand is assigned to public and private transport networks.

Each of these steps are applied separately for specific trip purposes. The conceptual flowchart defining computational procedure is given in Figure 1.2.

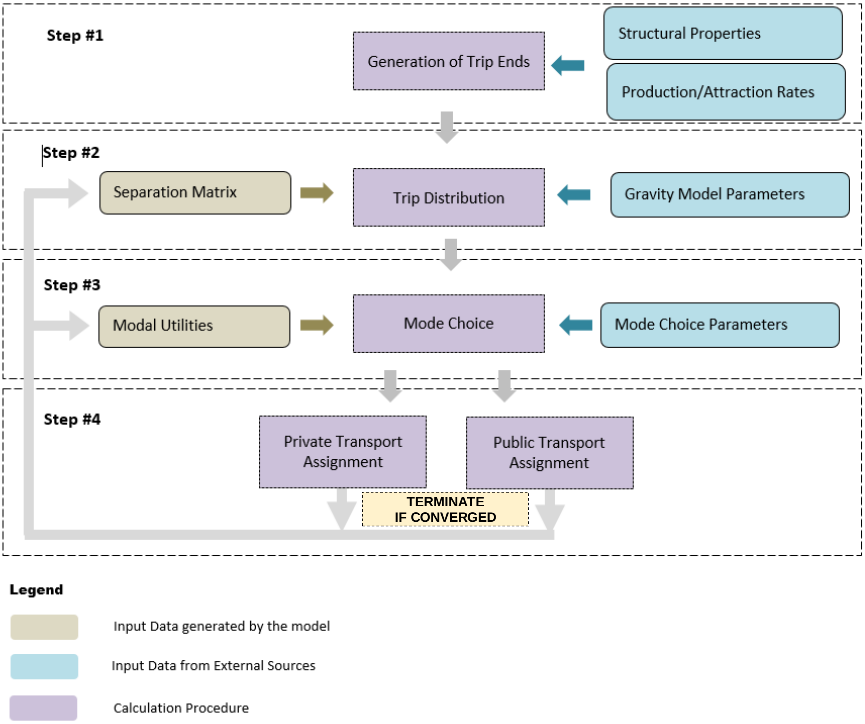

Fig. 1.2. Trip-based Travel Demand Modelling Procedure (Created by Author)

The procedure itself is of an iterative nature and the output of one step flows as an input into another step. Iterations terminate once some predefined convergence criterion has been reached (i.e. stability in the estimated demand matrices, flows on the network elements or travel times between origin and destination zones). A more detailed discussion of each step is given in the next sections that follow.

1.2.2. Step 1: Trip Generation

The basic assumption of the trip generation step is that the number of trips produced by, or attracted to, each zone can be estimated by analysing specific land use data. The trip generation procedure consists of two sub models: trip production and trip attraction.

While the trip production model estimates the number of trips O_i generated by each i_{th} zone in the study area, the trip attraction model has the role of estimating the amount of these trips D_j that will be attracted to each j_{th}zone of the study area. For instance, zones where jobs are located will mostly attract work related trips, whereas school (education) trips will be attracted to the zones where schools are located. As a matter of fact, conventional travel demand model treats productions of O_i and attractions D_j as corresponding to the actual trips or just as a potential/attractiveness for the trips to start or end at zone i. Generally, trip production or attraction model can be expressed as follows:

$$Q^k(i) = \sum_g SG_g(i) \cdot \alpha_g^k, \tag{1.1}$$

here $Q^k(i)$ – number of *k* purpose trips starting/ending at zone *i*; $SG_g(i)$ – intensity of structural property g at zone *i*; α_g^k – the production/attraction rate showing the number of *k* purpose produced/attracted trips per structural property g unit.

Structural property is a zone's attributes related to its land use, it can relate to the population size, area of commercial, recreational or other developments. Production or attraction rate expresses the number of generated or attracted *k* purpose trips per one structural property unit, e.g. the number of work trips per one resident. Average production and attraction trip rates are computed for each purpose from the observational data.

Outputs from each step of conventional demand model are inputs for the next step. As a result of the first step, productions and attractions are the inputs for the trip distribution model.

1.2.3. Step 2: Trip Distribution

The trip distribution model connects trip ends estimated in the trip generation model to determine trip interchanges between each zonal pair. This model also considers the effects of spatial separation and involves the application of the well-known Gravity Model.

The effect of spatial separation is usually different for different trip purposes, and as an example, it has been noted that in practice people are willing to travel to work relatively longer distances compared to other purposes.

Even though there is a large family of distribution models, the most commonly used is the gravity model. This model was originally generated from an analogy with Newton's gravitational law. Interestingly, the gravity model has been criticized for its rather loose derivation: why on earth should human behaviour necessarily comply with the same principles as gravitational bodies? Fortunately, Alan Wilson (1967) and later on several other scientists, developed a sound statistical theory underlying gravity models.

The gravity model states that the number of trips between an origin and destination pair is directly proportional to the number of productions at the origin and the number of attractions at the destination and inversely proportional to the spatial separation between zones. The formulation of the gravity model is as follows:

$$T_{ij} = A_i \cdot O_i \cdot B_j \cdot D_j \cdot f(c_{ij}), \tag{1.2}$$

here T_{ij} – number of trips originating at zone i and terminating at zone j; O_i – number of trips originating at zone i; D_j – number of trips terminating at zone j; f – separation function; c_{ij} – separation function argument defining separation between zones i and j; A_i and B_j – balancing factors for origin and destination zones, respectively.

Balancing factors A_i and B_j required to ensure the row and columns total constrains are calculated as follows:

$$A_i = 1 / \sum_j B_j \cdot D_j \cdot f(c_{ij}); \tag{1.3}$$

$$B_j = 1 / \sum_j A_i \cdot O_i \cdot f(c_{ij}). \tag{1.4}$$

The balancing factors A_i and B_j are, therefore, interdependent; this means that the calculation of one set requires the values of the other set. This suggests an iterative process, which works well in practice: given set of values for the separation/deterrence function $f(c_{ij})$, start with all B_j = 1, solve for A_i and then use these values to re-estimate all B_j; repeat until convergence is achieved.

Popular versions of the deterrence function, which represents the disincentive to travel are exponential (1.5), power (1.6) and combined (1.7) functions given below:

$$f(c_{ij}) = \exp\left(-\beta \cdot c_{ij}\right); \tag{1.5}$$

$$f(c_{ij}) = c_{ij}{}^{n}; \tag{1.6}$$

$$f(c_{ij}) = c_{ij}{}^{n} \cdot \exp\left(-\beta \cdot c_{ij}\right), \tag{1.7}$$

here c_{ij} – separation measure, defined by e.g. travel time between zones i and j; β, n – calibration parameters.

As we have seen, the parameters A_i and B_j are estimated as part of the iterative balancing algorithm. The parameters of the deterrence function must be calibrated to make sure that the observed trip lengths are reproduced as closely as possible. A naive approach to this task is simply to 'guess' or to 'borrow' a value for parameters, run the gravity model and then extract the modelled trip lengths. This should be compared with the observed trip lengths. If they are not sufficiently close, a new guess for parameters can be used and the process repeated until a satisfactory fit is achieved.

1.2.4. Step 3: Mode Choice

Models used to predict the usage of the available transport modes have their roots in the random utility theory that was developed by American econometrician Daniel McFadden who shared the 2000 Nobel Memorial Prize in Economic Sciences.

The most prominent and practically widely used Logit model was also derived by McFadden (1974) and this model is briefly presented and discussed in this section. Presentation is based on the work done by Train (1993).

A decision maker, labelled *n*, is faced with a choice among *J* alternatives. All possible alternatives are mutually exclusive, collectively exhaustive and comprise a finite choice set. It is assumed that the decision maker would obtain a certain level of benefit/utility from each alternative. The utility that decision maker *n* obtains from alternative *j* is U_{nj} for $j = 1, \ldots, J$. This utility is known to the decision maker but not by the researcher. The decision maker is assumed to be rational agent (so called perfect optimizer) and, therefore, chooses the alternative that provides the greatest benefit/utility. More strictly, the behavioural model is the following: choose alternative *i* if and only if $U_{ni} > U_{nj}\ \forall j \neq i$. Unfortunately, the researcher does not observe the decision maker's utility U_{ni}. The researcher observes some attributes of the alternatives as faced by the decision maker, labelled $x_{nj}\ \forall j$, and some attributes of the decision maker, labelled s_n and can specify a function that relates these observed factors to the decision maker's utility. The most common representation for U_{nj} is inspired by linear regression. The utility is separated into two additive parts: $U_{nj} = V_{nj} + \varepsilon_{nj}\ \forall j$. Where V_{nj} is called the deterministic or systematic part of the utility and ε_{nj} is the error term.

The researcher does not know $\varepsilon_{nj}\ \forall j$ and therefore treats these terms as random. The joint density of the random vector $\varepsilon'_n = \langle \varepsilon_{n1}, \ldots, \varepsilon_{nJ} \rangle$ is denoted as $f(\varepsilon_n)$. With this density, the researcher can make probabilistic statements about the decision maker's choice. The probability that decision maker *n* chooses alternative *i* is:

$$\begin{gathered} P_{ni} = Prob\left(U_{ni} > U_{nj} \ \forall j \neq i \right) = \\ Prob\left(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \ \forall j \neq i \right) = \\ Prob\left(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj} \ \forall j \neq i \right). \end{gathered} \tag{1.8}$$

In the Logit derivation, it is assumed that ε_{nj} terms are distributed independently, identically extreme value (IID EV) with probability density and cumulative density functions given in the formulas (1.9) and (1.10) respectively:

$$f(\varepsilon_{nj}) = -e^{-\varepsilon_{nj}} \cdot e^{-e^{-\varepsilon_{nj}}}; \tag{1.9}$$

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}. \tag{1.10}$$

Having in mind that each error term ε_{nj} is independent and treating ε_{ni} as given, we can express the probability that the decision maker *n* chooses alternative *i* as follows:

$$P_{ni|\varepsilon_{ni}} = \prod_{j \neq i} e^{-e^{-\left(\varepsilon_{ni} + V_{ni} - V_{nj} \right)}}. \tag{1.11}$$

Of course, ε_{ni} is not given and the choice probability is the integral of $P_{ni|\varepsilon_{ni}}$ over all values of ε_{ni} weighted by its density:

$$P_{ni} = \int \left(\prod_{j \neq i} e^{-e^{-\left(\varepsilon_{ni} + V_{ni} - V_{nj} \right)}} \right) e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}} d\varepsilon_{ni}. \tag{1.12}$$

Some algebraic manipulation of this integral results in a succinct, closed-form expression of Multinomial Logit model:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}. \tag{1.13}$$

Even though this choice idea is presented with reference to the individual decision maker, it is equally applicable to aggregate zone-based formulations.

One of the most prominent properties of the Multinomial Logit model is the property of independence from irrelevant alternatives (IIA). For any two alternatives *i* and *k*, the ratio of the Logit probabilities is as follows:

$$\frac{P_{ni}}{P_{nk}} = \frac{\dfrac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}}{\dfrac{e^{V_{nk}}}{\sum_j e^{V_{nj}}}} = \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni}-V_{nk}} . \tag{1.14}$$

The ratio does not depend on any alternatives other than *i* and *k*. That is, the relative odds of choosing *i* over *k* are the same no matter what other alternatives are available or what the attributes of the other alternatives are. The IIA property of Multinomial Logit models is a limitation for some practical applications. This limitation is often illustrated by the red bus/blue bus paradox (Ben-Akiva *et al.* 1994).

Consider the famous “red-bus-blue-bus” problem. A traveller has a choice of going to work by car or taking a blue bus. For simplicity assume that that the representative utility V_{nj} of the two modes are the same, such that the choice probabilities are equal: $P_{car} = P_{blue\ bus} = 1/2$. In this case the ratio of probabilities is equal to one: $P_{car} / P_{blue\ bus} = 1$. Now suppose that a red bus is introduced, and the traveller considers the red bus to be exactly like the blue bus. The probability that the traveller will take the red bus is therefore the same as for the blue bus, so that the ratio of their probabilities is one: $P_{red\ bus} / P_{blue\ bus} = 1$. However, in the Multinomial Logit model the ratio $P_{car} / P_{blue\ bus}$ is the same whether or not another alternative exists, therefore it remains at one. The only probabilities for which $P_{car} / P_{blue\ bus} = 1$ and $P_{red\ bus} / P_{blue\ bus} = 1$ are $P_{car} = P_{blue\ bus} = P_{red\ bus} = 1/3$. In real life, however, we would expect the probability of taking a car to remain the same when a new bus is introduced, and we would also expect the original probability of taking bus to be split between the two buses after the second one is introduced.

This example is rather stark and unlikely to be encountered in the real world. However, the same kind of misprediction arises with Multinomial Logit models whenever the ratio of probabilities for two alternatives changes with the introduction or change of another alternative. For example, suppose a new transit mode is added that is similar to, but not exactly like, the existing modes, such as an express bus along a line that already has standard bus service. This new mode might be expected to reduce the probability of regular bus by a greater proportion than it reduces the probability of car, so that the ratio of probabilities for car and regular bus does not remain constant. The Multinomial Logit model would overpredict demand for the two bus modes in this situation. The Logit choice models are extensively used in trip-based modelling systems. In a trip-based model, market segments are defined by trip purpose and household demographic groups, and the model predicts the probability of each mode for each origin destination (OD) pair. The model then allocates the fraction of trips for each segment and OD pair to modes in proportion to their predicted probabilities. This is an aggregate prediction, which is then summed up over all market segments to form trip tables.

1.3. Tour-based Modelling Approach

1.3.1. Main Idea

The tour-based model, presented in this section, explicitly recognizes the causal relationship between people's daily activities and their resulting mobility.

The demand model consists of three logical steps: generation of activity sequences, generation of tours coupled with mode choice and assignment. The final third step entails demand assignment to transport private and public network procedures. The flowchart, involving computational steps and input data requirements, is given in Figure 1.3.

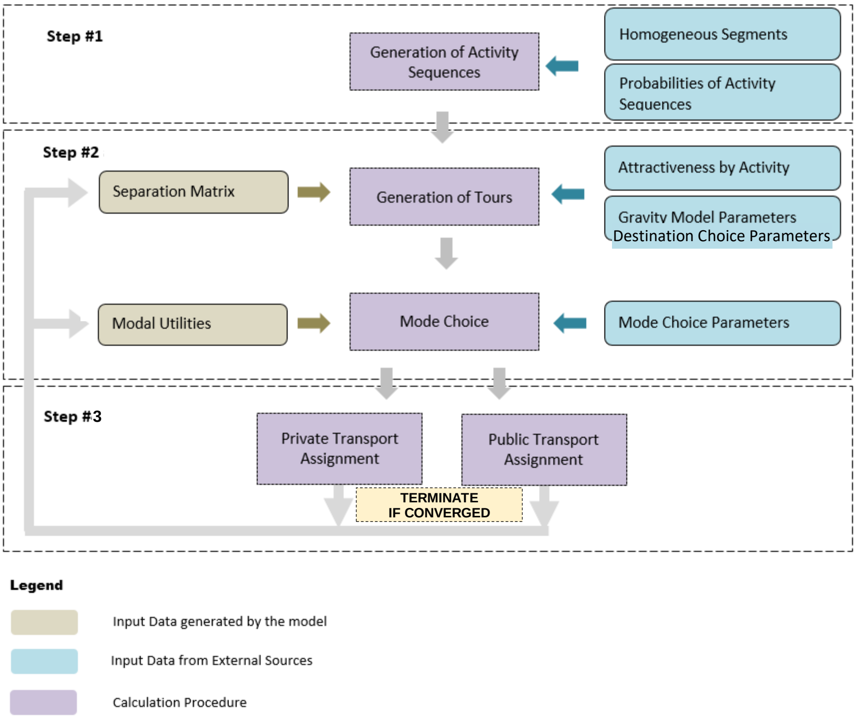

Fig. 1.3. Conceptual Representation of Tour-based Travel Demand Modelling Procedure (Based on Fellendorf *et al.* (2000))

The key input data within the first step are the probabilities of observed activity sequences distinguished by population segments. Probabilities represent the likelihood of the typical segment's member conducting the given activity sequence throughout the course of a typical weekday. Multiplication of these probabilities with the total segment's size results in the total number of activity sequences being carried out during the day. Next, destination choice and mode choice calculations are carried out in a single interdependent procedure. First, activity sequences are converted into location sequences (tours) by applying a gravity model and employing zonal attractiveness data. Then, mode choice, which distributes trips over available modes, is carried out considering the whole tour. Finally, the estimated demand is assigned to public and private transport networks. A more detailed description of each identified step is given in the following sections.

1.3.2. Step 1: Generation of Activity Sequences

To estimate activity sequences generated at each transport analysis zone, initially, the whole population needs to be classified into mutually exclusive and collectively exhaustive segments/groups, which feature homogeneous travel behaviour. The travel behaviour within segments needs to be as homogeneous as possible, whereas the behaviour between segments must be as different as possible. For example, a segmentation according to occupation and car availability criteria may look similar to that as given in Table 1.1.

Table 1.1. An Example of the Population Segmentation by Car Ownership and Occupation

Car Availability	Occupation			
	Employed	Unemployed	Student	Pupil
Available	Employed People with a Car (EC)	Unemployed People with a Car (NEC)	Students with a Car (SC)	Pupils with a Car (PC)
Not available	Employed People without a Car (ENC)	Unemployed People without a Car (NENC)	Students without a Car (SNC)	Pupils without a Car (PNC)

This classification, which was exemplified in literature (Schmiedel 1984; Fellendorf *et al.* 1997; Fellendorf *et al.* 2000) and is extensively used in PTV Visum software manual (PTV Group 2014), results into eight segments.

The classification can be accomplished according to any reasonable criterion; however, one needs to be quite considerate as the number of segments increases significantly with each additional criterion. And a vast amount of population segments might result into unreasonable requirements for sample size of the travel survey. And financial constraints can sometimes force the analyst into treating the whole population as single segment.

After identification of the population segments, activity sequences need to be identified and their probabilities need to be estimated, i.e. Table 1.2 is compiled. Probability represents the likelihood of the typical segment's member conducting the given activity sequence throughout the course of a modelled day (Fellendorf *et al.* 1997; Fellendorf *et al.* 2000).

Table 1.2. Probabilities of Activity Sequences

Activity Sequence	Population Segments				
	EC	ENC	NEC	NENC	…
MWM	P (MWM | EC)	P (MWM | ENC)	P (MWM | NEC)	P (MWM | NENC)	…
MDWM	P (MDWM | EC)	P (MDWM | ENC)	P (MKWM | NEC)	P (MKWM | NENC)	…
MWSM	P (MWSM | EC)	P (MWSM | ENC)	P (MWSM | NEC)	P (MWSM | NENC)	…
….	…	…	…	…	…

Abbreviations: M – Home, W – Work, D – Day Care Centre, S – Shopping Centre.

In general, the notation P(XYZ | S) expresses probability that a person will conduct a sequence XYZ during the day, given that that person belongs to segment S. These probabilities can be identified through the surveys by asking participating people about their activities conducted over a 24-hour period with all the relevant attributes such as their spatial locations and time references. Information revealing characteristics of trips made between activities (chosen modes, travel times etc.) is also collected as part of the survey.

Multiplication of each probability with the total number of people falling into the segment gives a total number of activity sequences conducted during the day by the members of that population segment.

The model developed by Fellendorf *et al.* (2000) does not include the time of day choice model for each activity that would use for example a discrete choice approach. Instead the time is modelled in the simplified fashion by defining daily patterns for the transitions between a pair of activities. More precisely, each activity pair, which signifies the movement between two activities, is allocated a probability distribution over the intervals within the modelled period. The probabilities define the proportions of movements (activity pairs) that are undertaken over the intervals within the modelled periods.

1.3.3. Step 2: Generation of Tours and Estimation of Mode Choice

Once the quantification of activity sequences is complete, the spatial location can be assigned to each activity within an activity sequence. Assignment of spatial locations is carried out sequentially, i.e. the first spatial location is assigned to the first activity, then conditional on the location of the first activity, destination choice is carried for the second activity and so on and so forth. This step is a significant improvement over trip-based approach as it explicitly considers the interdependence between trips within the trip sequence. This procedure utilizes the following information:

a) the activity specific attractiveness for each potential destination zone;
b) spatial separation between origin zone and potential destination zones expressed by travel times, distances, monetary costs or a combination of them;
c) the utility function parameters that defines the impact of various spatial separation variables.

The tour generation procedure usually relies on random utility theory and multinomial Logit model approach. If there are *k* spatial separation variables, the systematic part of utility between origin zone *i* and destination zone *j* is expressed as follows:

$$V_{ij} = R_{ij} + \ln\left(S_j\right); \tag{1.15}$$

$$R_{ij} = \sum_{k=1}^{K} \beta_k \cdot x_{ijk}, \tag{1.16}$$

here R_{ij} – spatial separation related systematic utility between i and j; x_{ijk} – values of *k* spatial separation variables (travel time, distance, cost, etc.); β_k – spatial separation utility function parameters/marginal utility of each spatial separation variable; S_j – activity's intensity at the zone *j*.

Intensity S_j (or size term) measures the activity opportunities at each potential destination. The size term depending on the activity being modelled is typically defined by various employment categories. For example, work activity is usually modelled as being attracted by the total employment, whereas shopping activity is assumed to be attracted by retail employment only.

Conditional on the current location i, probability of the next activity in the activity sequence being carried out at zone j is calculated by employing Logit model (Fellendorf *et al.* 2000):

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_{m=1}^{B} e^{V_{im}}} = \frac{e^{R_{ij} + \ln(S_j)}}{\sum_{m=1}^{B} e^{R_{im} + \ln(S_m)}} = \frac{S_j \cdot e^{R_{ij}}}{\sum_{m=1}^{B} S_m \cdot e^{R_{im}}}, \tag{1.17}$$

here P_{ij} – probability to choose j as a destination zone for the next activity in activity sequence, given current location at origin zone i; B – set of zones considered as potential destinations for the next activity in activity sequence.

Parameters β_k generally varies with population segment and activity being modelled. Lower β_k values result into the smaller destination choice sensitivity to separation (distance, travel time, etc), i.e. the attractiveness plays a major role in the destination choice. According to Fellendorf *et al.* (1997; 2000), empirical evidence shows that persons with a car available cover longer distances than persons without cars, which means that α parameters of population segments using a car have to be lower than those specified for groups without cars. Different activities also make difference, for example, α parameter specified for the activity 'Work' must be lower than that for activity 'Shopping' as urban population tend to make longer work-related trips and a rather short shopping-related trips.

In parallel to the destination choice, mode choice calculation is being undertaken. Precisely, once the destinations for first activity are located, the mode choice is calculated before progressing to the second activity. The probability of choosing mode t is estimated with a Multinomial Logit model.

$$P_{ijt} = \frac{e^{IMP_{ijt}}}{\sum_{h=1}^{H} e^{IMP_{ijh}}}, \tag{1.18}$$

here P_{ijt} – probability of choosing mode t for the trip between origin zone i and destination zone j; IMP_{ijt} – systematic utility of mode t between origin zone i and destination zone j; H – set of available modes.

Travellers' choices of destination and mode are importantly related. In the four-step model, mode choice was originally formulated as conditional on destination choice, being applied as the third step after the trip distribution, which has meant that destination choice is independent from the available modes. It can be argued that, in reality, destination choice inevitably relies on available modes to potential destinations. Modal service quality between origin zone i and destination zone j defines the zone j accessibility and, thus, is important during destination choice decisions. For example, some destination zones might be poorly accessed with a car, but are very easily reached by bus.

According to Transportation Research Board (2018), efforts to link mode and destination choice have most commonly involved the use of mode choice logsums as a multimodal impedance variable in destination choice models. In theory, if properly specified, this approach results in the destination-mode choice model system taking the mathematical form of a Nested Logit model. In the context of tour generation, multimodal impedance can be assigned to spatial separation related systematic utility R_{ij} to consider modal accessibility of the potential destinations:

$$R_{ij} = ln \sum_{t=1}^{T} e^{IMP_{ijt}} \quad , \tag{1.19}$$

here T – the number of available modes between origin zone i and destination zone j.

In this destination/mode choice context, the logsum term represents the maximum expected systematic utility that may be derived from the mode choice alternatives for each origin destination pair.

1.4. Travel Demand Assignment to the Transport Network

Assignment procedures form the core of any comprehensive transportation model. Procedures that allow for modelling a person's travel behaviour on their journey from an origin to a destination distribute a given travel demand to a network. This results not only in traffic volumes for road links and transit lines but also in indicators describing the service quality of a network (Friedrich *et al.* 2001).

The assignment model predicts the routes/connections that users will choose and the way that demand interacts with the available capacity (UK Department for Transport 2014a). According to Ortuzar *et al.* (2011) assignment procedures allow the obtaining of good aggregate network measures (e.g. total vehicle or passenger flows), to estimate zone-to-zone travel times for a given level of demand, to estimate the routes used between each origin and destination pair, to analyse which and to what extent origin-destination pairs use a link.

There are several various assignment procedures, which are used for specific purposes and in specific situations. In general, the assignment procedures are classified into two main groups: static and dynamic assignment procedures. The difference is that dynamic assignment procedures consider time-wise traffic variations, i.e. travel demand and network supply can be modelled as a time dependant phenomenon. Bearing in mind that the model being developed will entail an hourly travel demand resolution, in the following two sections only two specific dynamic assignment methods, which will be practically applied, are presented, namely dynamic stochastic private transport assignment and dynamic timetable-based public transport assignment.

1.4.1. Dynamic Stochastic Private Transport Assignment

The quantitative analysis of road network traffic performed through static assignment models yields the transport demand-supply equilibrium under the assumption that the state is constant over assignment period. This implies that the relevant variables of the system (i.e. flows, travel times, costs) are also assumed to be constant over the assignment period. Although static assignment models satisfactorily reproduce congestion effects on traffic flow and cost pattern, they do not allow for the reproduction of some important dynamic phenomena such as demand variation over time and of the time-varying supply (Bellei *et al.* 2005). For the use of these cases, dynamic assignment methods are necessary.

In the context of this thesis, the simplest PTV Visum dynamic stochastic private transport assignment method is presented and discussed. A concise high-level flowchart defining the method is given in Figure 1.4.

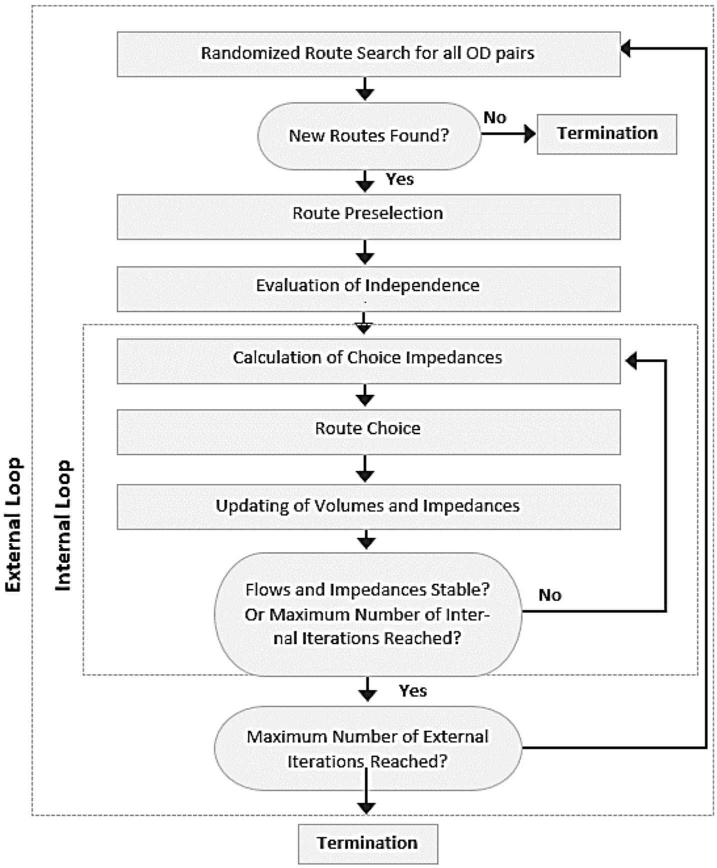

Fig. 1.4. Dynamic Stochastic Assignment Flowchart (Adopted from PTV Group (2014))

The algorithm is implemented with the two loops: external iteration loop for the route/connection search; and, internal iteration loop for the route/connection choice and network loading.

The procedure starts with a randomized route search, during which several alternative shortest path searches are undertaken by changing initial impedances of the network objects. They are drawn randomly from a normal distribution, which has the actual impedance as mean value and whose standard deviation σ is calculated as a function of R.

During the preselection stage newly found routes are checked as to whether they are of a significantly lower quality and can thus be deleted. Routes, whose travel time or impedance is significantly higher than that of the best route, are discarded.

All origin-destination pair routes that are found are then compared to each other to define their independence. The correction factor (CR_r) is calculated for each route according to the methods established by either Cascetta *et al.* (1996) or Ben-Akiva *et al.* (1999).

After establishing the routes and their independence, the procedure enters the internal loop for demand distribution, within which impedances are updated in the first place. Further the demand is distributed over the alternative routes and alternative departure times.

The outputs of the route choice step are traffic flows and separation impedances on the network elements (links and turns). Their fluctuation compared to the previous iteration is checked to see whether the stability is achieved. If they are stable, the algorithm proceeds with the route search, otherwise another internal iteration is undertaken.

Generally, the whole procedure stops/converges when the network is loaded with demand and no other alternative and viable routes can be found.

1.4.2. Timetable-based Public Transport Assignment

A public transport assignment problem requires a slightly more modified approach than a private transport assignment due to a more complex travel behaviour and a different network structure.

Car drivers may depart at any time and are free to choose a route which appears convenient to them. The travel behaviour of transit passengers, on the other hand, is strongly restricted by the line network and the timetable. Assignment procedures for a public transport network need to reflect the constraints imposed by line routes and timetables. They require specific search algorithms considering transfers between transit lines with their precise transfer time (Friedrich *et al.* 2001).

Within this section a timetable-based public transport assignment algorithm developed by (Friedrich *et al.* 2001; Friedrich *et al.* 2002; Friedrich 2008) and implemented in PTV Visum modelling package is introduced and discussed. This method takes the precise timetable into account,thus ensuring very precise results of the indicator data calculation. A concise high-level flowchart defining the method is given in Figure 1.5.

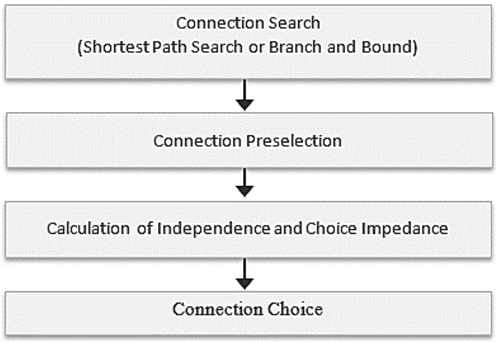

Fig. 1.5. Timetable-based Public Transport Assignment Flowchart (PTV Group 2017)

In its basic form, public transport assignment consists of four main steps: connection search, connection preselection, calculation of independence and choice impedance and connection choice.

The objective of the connection search step is to find all potentially attractive connections and calculate their impedance. The PTV Visum software implements connection search step in two different ways, either with shortest path search or branch and bound search. The fundamentals of the first method were first proposed by Dijkstra (1959) whereas Land *et al.* (1960) published the first description of the second method.

Within the public transport assignment context, during the shortest path search, one connection is identified for each origin-destination (OD) pair and each departure time. During the branch and bound search, however, not only the best connection is found for each OD pair and each departure time, but several good connections. This is achieved by building a single connection tree with a root node at the origin zone and other nodes located at alighting and boarding stops. The tree's edges are made up of connection segments representing either a walk or a transfer-free ride on one transit line. As the branch and bound approach reduces computing times and generates a richer connection choice set, it is regarded as superior over the shortest path algorithm and will be concisely described below. A detailed description of the method can be found in the paper written by Friedrich *et al.* (2002).

Branch and bound itself consists of two main procedures: the pre-processing step and the connection tree building step. In the pre-processing step route, segments are created which basically describe all potential movement alternatives across all single public transport lines. An example of route segments resulting on a single line running via four stops is given in Figure 1.6.

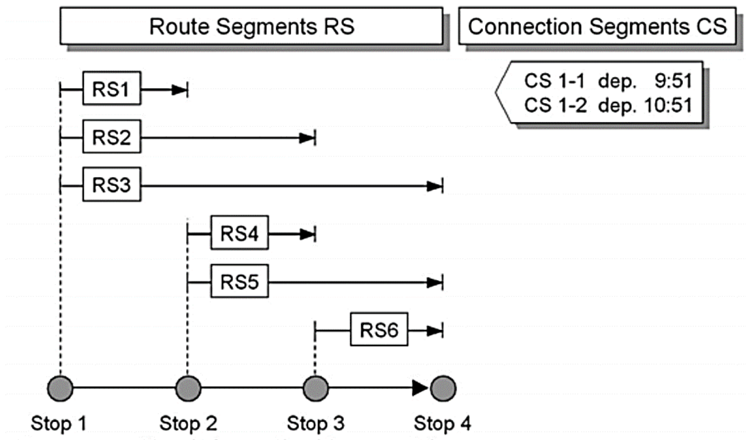

Fig. 1.6. Branch and Bound Pre-Processing Step (Friedrich 2008)

Route segments are converted to connection segments after amending spatial movement with the exact departure times. These connection segments are concatenated to connections later in the path building process. Figure 1.7 illustrates schematic representation of the connection tree. The root of the tree is the centroid node of the origin zone and the only outgoing branch represents the access (walk) from the origin to the first public transport stop. The following branches are concatenated to the first one if their temporal position is suitable, i.e. their departure time is later than the arrival time of the previous branch. In addition, there are other rules that bound the tree:

a) the connection segment will be added if and only if there are no other connection leading to the same stop point (node) that is by all means better i.e. it departs later, arrives earlier, has lower impedance and less number of transfers;

b) the connection segment will be added if and only if the resulting connection is not significantly worse that the best connection to the same stop point found so far.

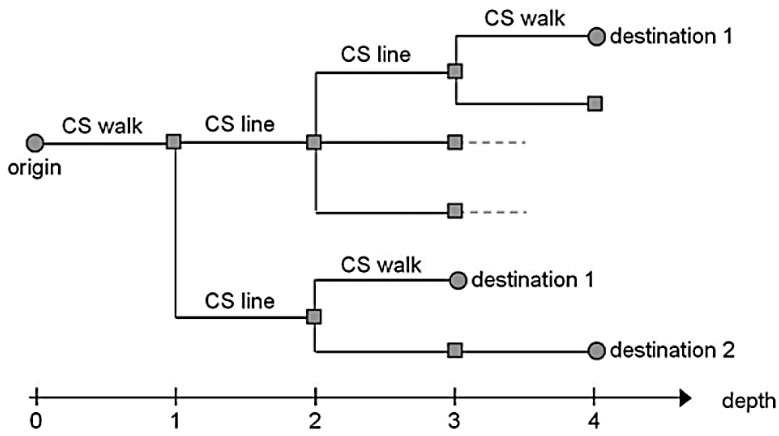

Fig. 1.7. Branch and Bound Connection Tree Building Step (Friedrich 2008)

The preselection step within the PTV Visum software compares and evaluates all found connections. Illogical connections unlikely to be used by passengers are discarded. Only convenient connections are offered to the passengers for the connection choice. That is, connection c_i between a particular origin destination pair is not deleted if all conditions are satisfied:

$$IMP(c_i) \le b_1 \cdot min_{c \in C}\ IMP(c) + b_2; \tag{1.20}$$

$$JT(c_i) \le d_1 \cdot min_{c \in C}\ JT(c) + d_2; \tag{1.21}$$

$$NT(c_i) \le e_1 \cdot min_{c \in C}\ NTR(c) + e_2, \tag{1.22}$$

here $IMP(c)$ – search impedance of connection c; $JT(c)$ – journey time of connection c; $NTR(c)$ – number of transfers of connection c; $b_1,\ b_2,\ d_1,\ d_2,$ $e_1,\ e_2$ – user defined constants; C – the set of all available connections between OD zones.

Each available connection is compared to the best one and removed from the memory if it is significantly worse regarding search impedance, journey time and number of transfers. The connection c_i search impedance mentioned above is measured in minutes and is a linear combination of journey time and number of transfers; its components are following:

$$\begin{aligned} IMP(c) = {} & \alpha_{IVT} \cdot IVT(c) \cdot \alpha_{ATR} \cdot \mathrm{ATR}(c) \cdot \alpha_{AT} \cdot \mathrm{AT}(c) \cdot \alpha_{ET} \cdot \mathrm{ET}(c) \cdot \\ & \alpha_{WKT} \cdot \mathrm{WKT}(c) \cdot \alpha_{TWT} \cdot \mathrm{TWT}(c) \cdot \alpha_{NTR} \cdot \mathrm{NTR}(c), \end{aligned} \tag{1.23}$$

here $IVT(c)$ – in vehicle time; $ATR(c)$ – auxiliary transport ride time; $AT(c)$ – access time; $ET(c)$ – egress time; $WKT(c)$ – walk time;

$TWT(c) -$ transfer wait time; $NTR(c) -$ number of transfers; $\alpha -$ weighting factor.

The final assignment step models the passengers' choice among the remaining set of connections. However, before this step, the choice impedance (a separate measure than search impedance) and similarity of the connections need to be calculated. It is assumed that passengers' decisions to choose a particular connection are based on the perceived journey time(PJT), the difference between the desired and the actual departure time (ΔT_{early}, ΔT_{late}) and the fare. They are combined to a single variable describing connection choice impedance.

In PTV Visum, the choice impedance of a connection c_i used in the connection choice in a time interval a is calculated as follows:

$$IMP^{a}(c) = f_{PJT} \cdot PJT(c) + f_{Fare} \cdot FARE(c) + f_{\Delta T^{early}} \cdot \Delta T^{ae}(c) + f_{\Delta T^{late}} \cdot \Delta T^{al}(c); \tag{1.24}$$

$$PJT(c) = IVT(c) \cdot \beta_{ATR} \cdot \mathrm{ATR}(c) \cdot \beta_{AT} \cdot \mathrm{AT}(c) \cdot \beta_{ET} \cdot \mathrm{ET}(c) \cdot \beta_{WKT} \cdot \mathrm{WKT}(c) \cdot \beta_{TWT} \cdot \mathrm{TWT}(c) \cdot \beta_{NTR} \cdot \mathrm{NTR}(c). \tag{1.25}$$

The impedance of the path is measured in minutes and is a linear combination of journey time and number of transfers, its components are following:
$PJT(c)$ – perceived journey time of connection c_i; $FARE(c)$ – fare of connection c_i; $\Delta T^{ae}(c)$ – temporal displacement of connection c_i with respect to demand time interval a if connection is available earlier than the temporal location of demand interval a; $\Delta T^{al}(c)$ – temporal displacement of connection c_i with respect to demand time interval a if connection is available later than the temporal location of demand interrval a; β – weighting factor.

All perceived journey time components have been defined in search impedance expression – their meaning remains identical.

All distribution models, in their basic form, cannot consider interactions between different connections in a timetable-based assignment. However, ignoring this aspect can be a drawback. Therefore, analogously to the independence of private transport routes, independence of connections must be estimated. Friedrich *et al.* (2001; 2002) has proposed and implemented the following formulation of the correction factor $\lambda(c)$ and commonality factor $f_c(c)$:

$$\lambda(c) = \frac{1}{\sum_{c'} f_c(c')}; \tag{1.26}$$

$$f_c(c') = \left(1 - \frac{x_c(c')}{s_x}\right) \cdot \left(1 - \gamma \cdot min\left(1, \frac{s_z \cdot |y_c(c')| + s_y \cdot |z_c(c')|}{s_y \cdot s_z}\right)\right); \tag{1.27}$$

$$x_c(c') = \left(\frac{\left|DEP(c) - DEP(c')\right| + \left|ARR(c) - ARR(c')\right|}{2} \right); \tag{1.28}$$

$$y_c(c') = PJT(c) - PJT(c'); \tag{1.29}$$

$$z_c(c') = FARE(c) - FARE(c'), \tag{1.30}$$

here $\lambda(c)$ – correction factor; s_x, s_y, s_z – parameters controling the range of influence in temporal distance, difference in perceived journey time and difference in fare, respectively; γ – calibration constant; $x_c(c')$ – temporal distance between c and c'; $y_c(c')$ – the perceived journey time overhead of c' with respect to c; $z_c(c')$ – fare overhead of c' with respect to c; $DEP(c)$ – departure time of connection c; $ARR(c)$ – arrival time of connection c.

The use of connections' independence is highly advised by guidelines for public transport assignment methods. For example, Web TAG Unit M3.2 (UK Department for Transport 2014b) suggests that temporal proximity, perceived journey time and fare differences should be combined to derive an independence of connection factor, which defines the attractiveness of a particular connection relative to all others.

The share of the travel demand of time interval a is assigned to connection c proportionally to its utility (inverse of choice impedance). In the case of the standard Multinomial Logit model, the equation taking into account the correction factor $\lambda(c)$ is as follows:

$$P^a = \frac{e^{-\cdot\beta\cdot IMP^a(c)}\cdot\lambda(c)}{\sum_{c \grave{o} C} e^{-\cdot\beta\cdot IMP^a(c)}\cdot\lambda(c)}. \tag{1.31}$$

The presented approach is implemented in the PTV Visum transportation model and it allows a more realistic distribution of passengers' trips on the available public transport connections compared to the basic Multinomial Logit form.

1.5. A Note on Transport Modelling Environment

PTV Visum is the software for traffic analysis, forecasts and GIS-based data management. It consistently models all road users and their interactions and over time has become widely used in the field of transport planning.

The software developers of PTV Visum are PTV Group in Karlsruhe, Germany. The company was founded in 1979 by Dr. Hans Hubschneider at the Karlsruhe Institute of Technology in Germany. PTV Group currently employs over 700 engineers and scientists most of whom actively work in academia.

The transport model, implemented in a software product PTV Visum, consists of the following components (Friedrich 1998):

1. The demand model contains the travel demand data. It models activities and generates tours to estimate and forecast mode-specific origin-destination matrices for behaviourally homogeneous population groups. Travel demand estimates are derived from structural land use data and service indicators determined by the impact model. Because the demand model describes the traveller's choice between various modes of transport it may be labelled "multi-modal".
2. The network model contains the relevant data of the supply systems describing their spatial and temporal structure. It consists of traffic zones, nodes and public transport stops, links and public transport lines. Since the network model integrates private and public transport modes it may also be labelled as "multi-modal".
3. The impact models take their input data from the demand model and the network model. PTV Visum integrates different impact models to analyse and evaluate a transport system. A user model simulates the travel behaviour of public transport passengers and car drivers. Based on routing and assignment models it calculates traffic volumes and service indicators, e.g. travel time, number of transfers or service frequency. An operator model determines operational indicators of a public transport service, like vehicle kilometres, number of vehicles or operating cost. An environmental impact model provides several methods to assess the impacts of motorised transport on the environment.

Even though this particular software is a commercial product, the company supports master and PhD students with special academic free of charge licenses that can be applied for scientific purposes. This feature makes the product a very attractive tool not only for practitioners, but for the scientists as well. Thus, this modelling package with underlying algorithms has been selected as a primary environment for transport modelling related tasks.

1.6. Conclusions of the First Chapter and Formulation of the Objectives of the Thesis

1. Tour-based travel demand modelling technique was devised at the very start of the 21th century and later applied to several studies in Europe

(Hannover, Prague, St. Gallen etc.) and elsewhere (Izmit etc). Even though it features several improvements over the trip-based models, its wider spread is limited by technical, institutional and other reasons. In the context of national transport modelling state of the art this approach is innovative and has never been tested or implemented in practice.

2. The trip-based model's issue of treating a person's trips as independent decisions, where the effects of preceding trips on the consequent ones are not considered, can be resolved by tour-based models as they deal with the sequences of activities and trips (a.k.a. tours) and consider their interlocking dependencies. Therefore, it is concluded that tour-based travel demand modelling is an appropriate tool that can be applied in a national and international modelling practice to replace outdated trip-based aproach until agent-based models achieve their mature stage.
3. On the other hand, it is worth recognizing that the tour-based approach is only a temporary step forward as agent-based models (ABM) are the actual state of the art in the global context especially United States of America. Yet, the practical ABM application in the national context is restricted currently by the absence of readily available datasets (detailed private and public transport networks, disaggregate spatial distribution of the population, detailed travel behaviour patterns etc.), application complexity (software development skills, complex model structure, etc.) and as a result by a potentially very high development cost.
4. Due to the whole day supply and demand modelling, the assignment methods of tour-based travel demand models must consider time-wise traffic variations, and therefore must be dynamic. Dynamic stochastic private transport assignment and timetable-based public transport assignment are deemed to be the most appropriate for tour-based travel demand model and will be applied in later model development stages.

Having reviewed travel demand model development techniques, the following objectives were set for the remainder of the thesis:

1. To develop and test efficient and up-to-date data collection techniques ensuring representation of travel supply and travel demand for the whole day.
2. To create a travel demand model representing daily demand and supply for planning of urban transport system and apply it in the assessment of the impact of the transport network development scenarios.
3. To develop a comprehensive methodological framework for urban transport modelling, which would guarantee statistically reliable models and could be easily applicable for planning of transport networks.

2

Travel Demand Empirical Research: Methodology, Data Collection and Statistical Analysis

This 2nd chapter forms a significant part of the thesis by means of empirical research of travel behaviour in Vilnius city. One purpose of this empirical research is to better understand and quantify several mobility parameters such as:

1. Proportion of Travelling Respondents. This represents the share of the sampled respondents who have chosen to travel on the reported day. The proportion of travelling respondents can also be used to approximate a probability of the respondent's choice to travel on any given weekday.
2. Average Number of Trips. This represents the mean number of trips being undertaken by a random individual.
3. Mode Split. This represents the relative frequencies of modes that were chosen to make the trips within the area of interest.
4. Distribution of Trip Purposes. This represents the relative frequencies of activities that were undertaken at the destinations of all the trips made by sampled residents.
5. Average Trip Length. This represents the average distance of the trips that were undertaken by sampled residents.

6. Activity Start Times. This represents the relative frequencies of the start times of the observed activities.
7. Daily Activity Sequences and their probabilities. Activity sequences represent the activities being undertaken over the course of the typical weekday, whereas probabilities represent the likelihoods of those sequences being conducted on the reported day.

All these parameters play a significant role in the development, calibration and quality assessment of a typical tour-based travel demand model.

This empirical research is based on a travel diary survey that was planned and executed in cooperation with the Municipality Enterprise "Vilniaus Planas" in 2017. The company funded the administration of the survey as part of its ongoing development of Vilnius Sustainable Urban Mobility Plan (Vilniaus Planas …. 2018). The following sections describe in more detail the research object, the questionnaire design, sampling strategy and the results.

There have been 2 scientific articles (Dumbliauskas and Barauskas 2015; Dumbliauskas *et al.* 2017b) published on the topic of this chapter.

2.1. Study Area Description

Vilnius is the capital of Lithuania and its largest city, with a population of 533,000 residents as per the most recent census data (The Department of Statistics 2011). In this empirical research, the focus is placed on the administrative area of Vilnius City Municipality, for which a detailed definition of its spatial extents is given in Figure 2.1.

The area of Vilnius City covers 403 km^2 in total and bearing in mind the number of residents reported above, it has a density of approximately 1,320 residents per square kilometre. The following illustration in Figure 2.1 also contains a more detailed visualisation of residential density within the administrative boundaries.

The area of interest was firstly divided into 42 primary transport analysis (TAZ) zones that in terms of spatial coverage agree with the administrative borough areas. Secondly, each primary zone was divided into several secondary TAZs. The spatial extent of the secondary TAZs was governed by the understanding that these zones will be maintained in the later stages of travel demand model development, and by the desire to maintain land use homogeneity and a reasonable model resolution with the size of each zone being close to 3,000 residents. This division resulted in 218 transport analysis zones with the average area being equal to 1.85 km^2 and the average population size being equal to 2,830 residents. The technical data management and visualisation was undertaken using the open source geographic information system QGIS.

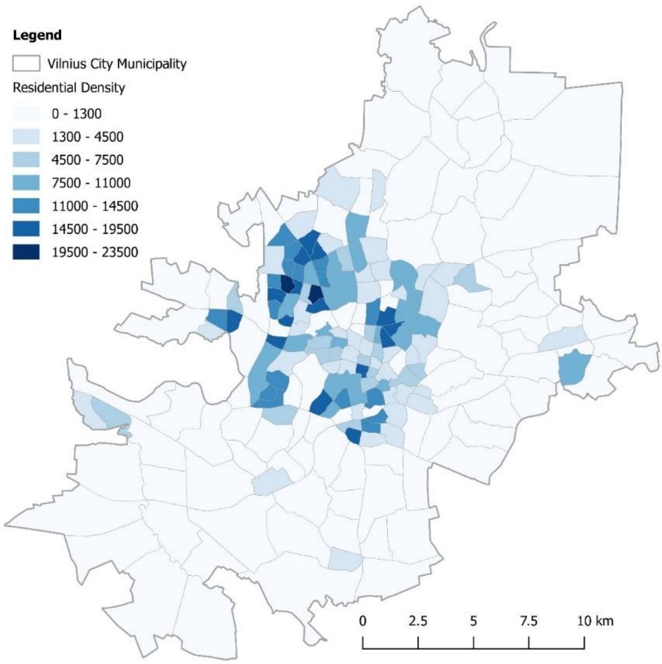

Fig. 2.1. A Map of a Study Area
(Created by Author using Data from The Department of Statistics 2011)

To identify travel behaviour patterns and mobility parameters within the depicted area, a travel diary survey has been developed and conducted. The next section describes the principles and processes that were followed in the questionnaire design stage.

2.2. Questionnaire Design

The writing of questionnaires for the data collection was an in-depth process; as much as an art as it is a science. There are a few widely accepted and well-known principles, which should be considered while compiling the questionnaire (Hernsher *et al.* 2005):

1. Appropriateness. The necessity of every question should be justified by its relevance to the hypothesis being tested. "Nice to know" questions add an unnecessary complexity and should be avoided.

2. Simplicity. The questionnaire should use simple language and unambiguous wording, otherwise there is a risk for confusion and misunderstanding.
3. Avoidance of leading (loaded) questions. The wording of a question must be as neutral as possible, without any hidden suggestions. If we ask such leading questions, decision makers are then compelled to answer in a manner that they might not necessarily support simply for fear of ostracism should they answer in the manner that best represents their true position.

The questionnaire was designed to provide information on travel behaviours for the tour-based travel demand modelling procedure. On a future basis, this kind of survey ideally should be repeated in Vilnius city about every five years to provide data for retrospective analysis and input for demand model update.

Germany can be highlighted as a good practice example, where it is recognized that up-to-date information about people's travel and mobility behaviour is indispensable for transportation policy decisions and planning. Only on the basis of such information the transportation infrastructure can be designed and preserved in order to meet the needs of the population – today and in the future. Since 1994 these German Mobility Panel surveys have been financed by, and carried out on behalf of the German Federal Ministry of Transport and Digital Infrastructure. This survey collects information about the household's travel behaviour over a seven-day period within three consecutive years. The Institute for Transport Studies of the Karlsruhe Institute of Technology is responsible for the design and scientific supervision of the survey (Institute for Transport Studies 2017).

Questionnaire was comprised of three main sections. The purpose of the first section was to familiarize the respondent with the relevant definitions such as an activity, trip and stage. A clear distinction between stage and trip is essential as people tend to report stages (continuous movement with one mode of transport/one vehicle) as complete trips, even though they connect only one activity with transfer between modes, rather than two activities.

Then, respondents reported the trips carried out during a recent weekday (Monday-Friday). It was specifically chosen to request the information for the most recent working day with a hope that this strategy will decrease cognitive and memory burden on the side of respondent and at the same time potentially increase response rate. For every trip made during the 24-hour period, respondents recorded the activity, origin, destination, modes used and the time of day.

Finally, respondents were asked to identify their main sociodemographic characteristics: gender, age, highest attained education level, occupation etc.

2.3. Sampling Strategy

The power of sampling is its ability to approximate from a small group the characteristics of the whole population within a known margin of error. The procedure usually consists of the following steps (Hernsher *et al.* 2005):

1. Determination of the sample frame.
2. Selection of the survey mode.
3. Identification of the sampling strategy.
4. Identification of the sample size.

The sampling frame represents the universal but finite set of decision makers to whom the analyst may administer the data collection instrument. In this thesis, the sample frame consists of the population living within the area of interest which is depicted in Figure 2.1. Due to legal and timeframe constraints, it was decided to reduce the sampling frame by excluding people younger than 16 years.

Generally, surveys may be accomplished in four survey modes: mail, phone, face-to-face or internet. It has been chosen to conduct this survey in two modes:

1. The first survey step was carried out over June and July months in 2017. Respondents were contacted and asked to fill in the survey form online.
2. The second survey step was conducted in September 2017. The respondents were visited at their household premises and interviewed face-to-face.

The respondents for the sample were chosen using multistage stratified sampling procedure that can be defined by the following stages:

1. First stage – each of 42 primary TAZ has been assigned the fraction of the sample proportional to the size of the population of that primary TAZ.
2. Second stage – the sample of each primary TAZ was distributed to secondary TAZs proportionally to the population of each secondary TAZ.
3. Third stage – respondents were chosen randomly from the population of the secondary TAZ. The sample's proportional distribution of age and gender has been maintained to be as close to the secondary TAZ's distribution as possible.

In practice, the sample size is defined by budgetary considerations most of the time and this has been the case and within the context of this work. Budgetary constraints allowed us to sample 1,773 respondents out of the whole population.

Survey results has been used to estimate various mobility parameters, such as proportion of travelling respondents, average number of trips, average trip length, mode share etc. Further chapters presenting the results report sample size, sample standard deviation and margin of error (95% confidence interval) for the proportion of travelling respondents and the average number of trips. However, due to space constraints, a confidence assessment has not been provided for the remaining mobility parameters.

2.4. Travel Survey Data Analysis

2.4.1. Data Analysis Tools and Methodology

Within the scope of getting valuable results from raw data, four main tools have been utilized:

1. Pandas Library (McKinney 2008).
2. Numpy Library (Oliphant 2006).
3. Geocoder Library (Carriere 2018).
4. Microsoft Excel spreadsheets.

Pandas is a Python library written for data manipulation and scientific analysis. It offers data structures and operations for manipulating numerical tables and time series. Moreover, it is free software released under the three-clause BSD (Berkeley Software Distribution) license (McKinney 2013).

Numpy is a Python library adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

Geocoder is a simple and consistent Python geocoding library allowing to establish a mapping between addresses and spatial location defined in latitude-longitude format.

The analysis has been conducted in two major steps:

1. Raw data cleaning, filling and arrangement into the format ready for analysis have been carried out using Pandas, Numpy and Geocoder libraries in Python language.
2. Data analysis and visualisation were carried out with Microsoft Excel spreadsheets.

The main distinctive feature of this analysis is that it presents most of the travel behaviour parameters by classifying the whole population into groups. Most of the parameters were classified by residents' characteristics that were observed during the survey e.g. age, gender, occupation and education. Statistical analysis starts off with an isolated look at the distributions of these respondents' characteristics and later delves deeper into the investigation of mobility parameters.

2.4.2. Respondents' Characteristics

The beginning of the analysis started with the distributions of the respondents among the levels of sociodemographic characteristics. Distribution across the age groups is given in Figure 2.2.

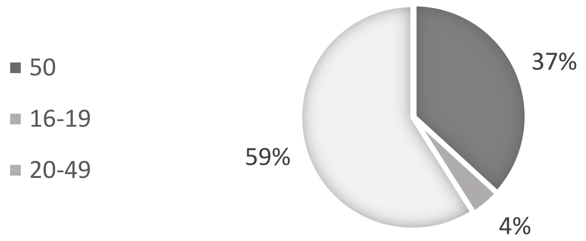

Fig. 2.2. Distribution of Age (Created by Author)

In summary, the total number of respondents surveyed is 1773. People aged between 20 and 49 comprised 59% of the sample and residents that were 50 or older accounted for 37% of the respondents. It is worth noting that no data has been collected for the youngest population category of between 6 and 16 years old. This is a major dataset drawback, which resulted out of legal constraint. It is recognised that this will present some limitations in the accuracy of the modelling results.

Distribution across gender categories is shown in Figure 2.3.

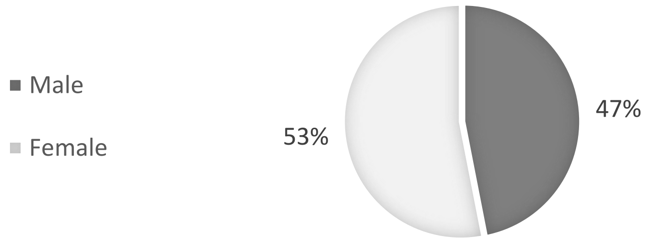

Fig. 2.3. Distribution of Gender (Created by Author)

The distribution among gender levels meets a prior expectations and remains compliant with the census data, which identifies slightly higher proportion of females.

Distribution across occupancy groups is given in Figure 2.4.

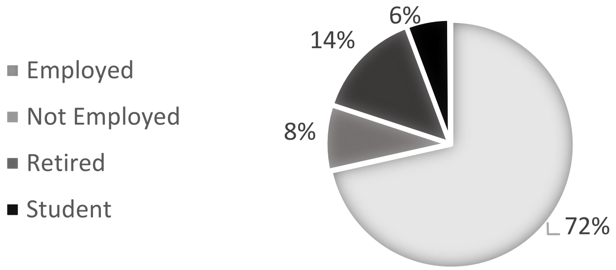

Fig. 2.4. Distribution of Occupancy (Created by Author)

In terms of occupation, it has been found that the urban population is highly economically active as employed people comprise 72% of the sample and students contribute with another 6%. Unemployed people account for 8% of the sample, whereas retired people account for the remaining 14%.

Distribution across education levels is outlined in Figure 2.5.

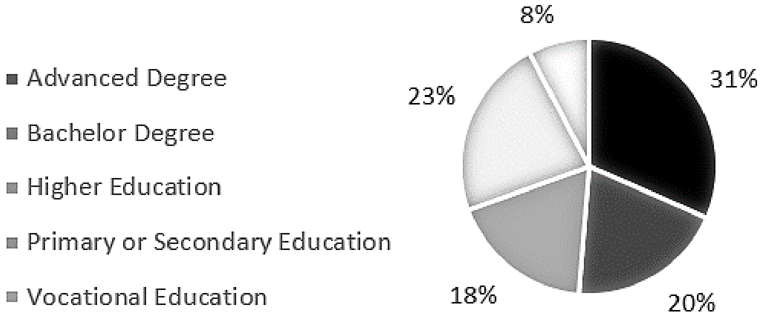

Fig. 2.5. Distribution of the Highest Attained Education Level (Created by Author)

It has also been found that the sampled population is quite highly educated. As an example, the total proportion of people possessing Bachelor, Master or PhD degrees is 51%.

At this point there is a natural expectation that urban travel behaviour supposed to be rather intense due to the high education and employability levels. This will be analysed in more depth in the following sections.

2.4.3. Identification and Analysis of Probability of Travelling

The analysis of the proportion of respondents who travelled (PT) on the reported day is presented in this section. Other related statistics, such as sample size, sample standard deviation and margin of error for 95% confidence interval are also calculated. The margin of error is calculated assuming that the proportion sampling distribution is nearly normal. Sample standard deviation was calculated using the following expression:

$$\sigma = \sqrt{\hat{p} \cdot (1 - \hat{p})}\,, \tag{2.1}$$

here $\hat{p}$ – sample proportion.

Margin of error formula:

$$m = t \cdot s\,, \tag{2.2}$$

here t – critical value; s – standard deviation of sampling distribution/standard error.

Standard error formula:

$$s = \frac{\sqrt{\hat{p} \cdot (1 - \hat{p})}}{\sqrt{n}}, \tag{2.3}$$

here $\hat{p}$ – sample proportion; n – sample size.

Critical value expression:

$$t = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \tag{2.4}$$

here Φ^{-1} – inverse cumulative distribution function of the Student distribution with *n–1* degrees of freedom; α – significance level (0.05).

A calculated margin of error provides an opportunity to identify the potential range within which a true value of the parameter might be.

It has been found that the overall share of travelling urban residents is equal to 0.8, which means that about 80% of the sample people were travelling on the reported day. Table 2.1 gives some more detailed insight into the behaviour of urban residents with the PT classified by several respondents' characteristics.

Table 2.1. Proportion of Travelling Respondents

Variable	Level	Proportion	Standard Deviation	Sample Size	Margin of Error
Age	50	0.7	0.48	652	0.04
	20–49	0.8	0.38	1046	0.02
	16–19	0.8	0.43	75	0.10
Gender	Male	0.8	0.41	832	0.03
	Female	0.7	0.44	941	0.03
Education	Primary or Secondary Education	0.6	0.49	404	0.05

End of Table 2.1

Variable	Level	Proportion	Standard Deviation	Sample Size	Margin of Error
	Vocational Education	0.7	0.47	137	0.08
	Bachelor's Degree	0.9	0.34	350	0.04
	Higher Education	0.7	0.46	322	0.05
	Advanced Degree	0.9	0.35	560	0.03
Occupa-tion	Unemployed	0.6	0.48	151	0.08
	Employed	0.8	0.36	1268	0.02
	Retired	0.4	0.49	253	0.06
	Student	0.7	0.45	99	0.09

From the table presented above, it can be defined that younger, more economically active and more educated people generally feature higher chances of leaving home on a given weekday. It is also worth noting the small difference in travel behaviour between male and female, with the male proportion being slightly higher. Next, it is interesting to explore the same parameter but categorised by age and education. Information is given in Table 2.2.

Table 2.2. Proportion of Travelling Respondents Categorised by Age and Occupation Levels

Age Levels	Occupation Levels	Proportion	Standard Deviation	Sample Size	Margin of Error
50 and over	Employed	0.8	0.38	365	0.04
50 and over	Unemployed	0.5	0.50	33	0.17
50 and over	Retired	0.4	0.49	253	0.06
16–19	Employed	0.7	0.45	21	0.19
16–19	Student	0.8	0.42	54	0.11
20–49	Employed	0.9	0.35	882	0.02
20–49	Unemployed	0.7	0.48	118	0.09
20–49	Student	0.7	0.47	45	0.14

It is interesting to observe that even the least active group (aged over 50 and retired) features a rather high proportion of travelling respondents equal to 0.4.

2.4.4. Identification and Analysis of Average Number of Trips

The analysis of average number of trips (ANT) is presented and discussed within the scope of this section. Other related statistics, such as sample size, sample standard deviation and margin of error for 95% confidence interval are also calculated. The margin of error is calculated assuming that the proportion sampling distribution is nearly normal:

$$\hat{\sigma} = \sqrt{\frac{\Sigma\left(x_i - \hat{x}\right)^2}{n-1}}, \tag{2.5}$$

here $\hat{x}$ – sample average; x_i – sample data point; n – sample size.

Critical value, standard error and margin of error computation method remains identical to the one defined in fomulas (2.2) to (2.4). A calculated margin of error provides an opportunity to identify the potential range within which a true value of the parameter might be.

The average number of trips in the overall sample is 2.3. As it was the case with the proportions of travelling respondents, the average number of trips also varies across the respondents with different characteristics and Table 2.3 allows a more thorough examination of ANT parameter.

Table 2.3. Average Number of Trips by Respondents' Characteristics

Variable	Level	Average	Standard Deviation	Sample Size	Margin of Error
Age	50 and over	1.8	1.7	652	0.13
	20–49	2.6	1.8	1046	0.11
	16–19	2.2	1.7	75	0.37
Gender	Male	2.4	1.8	832	0.12
	Female	2.2	1.7	941	0.11
Education	Primary or Secondary Education	1.6	1.5	404	0.14
	Vocational Education	1.9	1.7	137	0.29
	Higher Education	1.8	1.4	322	0.16
	Bachelor's Degree	3.0	1.8	350	0.19
	Advanced Degree	2.8	1.8	560	0.15
Occupation	Unemployed	1.8	1.67	151	0.27
	Employed	2.6	1.75	1268	0.10
	Retired	1.0	1.28	253	0.16
	Student	1.9	1.46	99	0.29

From the Table 2.3, it is obvious that all categorical variables have the power to explain the average number of trips. However, it is worth noting the small difference between males and females with the latter being slightly less active.

In terms of age, the most mobile are middle-aged (20–49) urban citizens with ANT equal to 2.6 whereas elders (50 years and over) make almost one trip less during any given weekday. The estimate for the youngest group is rather uncertain and the only conclusion can be made about the ANT parameter being somewhere between middle aged people and elders.

Further consideration of education, leads to the general trend: more educated people travel more, i.e. people having Bachelor or Advanced (Masters or PhD) degrees make 3.0 and 2.8 trips accordingly.

The ANT estimates for different occupations also varies considerably with the numbers meeting a priori expectation. The most mobile employed group makes 2.6 trips per day whereas retired people are the least mobile with ANT being equal to 1.0. Unemployed people and students have somewhat similar ANT values with their estimates being rather uncertain.

A further step is to explore how age and occupation together affect ANT: the analysis is presented in Table 2.4.

Table 2.4. Average Number of Trips by Age and Occupation

Age	Occupation	Average	Standard Deviation	Sample Size	Margin of Error
50 and over	Employed	2.4	1.7	365	0.18
50 and over	Unemployed	1.5	1.6	33	0.55
50 and over	Retired	1.0	1.3	253	0.16
16–19	Employed	2.3	1.8	21	0.79
16–19	Student	2.1	1.6	54	0.42
20–49	Employed	2.7	1.8	882	0.12
20–49	Unemployed	1.9	1.7	118	0.30
20–49	Student	1.6	1.3	45	0.37

In general, it is seen in the above tables that both factors affect the ANT: older and less economically active people travel less.

2.4.5. Identification and Analysis of Mode Split

Mode split (MS) has an interesting role of revealing how well developed and attractive various transport systems within the analysis area are. Within this section the overall as well as homogeneous group specific mode split statistics are presented and discussed.

The survey captured all used travel modes (stages) for each trip, however the duration or lengths of each mode within one trip has not been identified and, therefore, we proceed with an analysis that is based on the main modes. The main trip mode was assumed considering the following order of modal priority: car, public transport, motorbike, taxi, bike and walk. For example, if the car, public transport and walk modes are observed within a trip, the car receives a main mode label due to it being earlier in the modal priority list.

It is also important to note that the estimated modal shares given in Table 2.5 are based on two different measures: the number of trips and the estimated distance. Modal split based on first estimation method is more common among practitioners, because the distance is rarely obtained during the surveys. The second method considers each trip's distance and reveals the actual modal usage within the sample.

The total number of observed urban residents' trips were 4,049, therefore, the statistical reliability of mode shares is considered as very high.

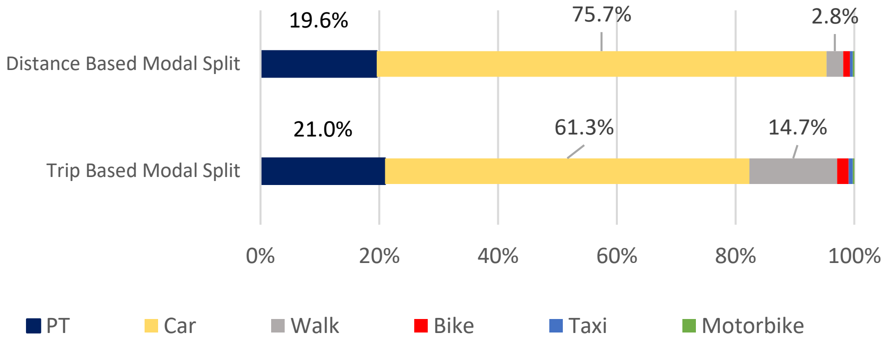

Fig. 2.6. Sample Modal Shares (Created by Author)

It is clear that the sampled persons rely on three main modes i.e. public transport, car and walk for daily mobility needs with the car mode being dominant. The trip-based assessment reveals that car share is equal to 61% whereas public transport with walk modes take up 21.0% and 14.7% respectively. Due to walking being used only for short nature related travel, the walk mode accounts for only 2.8% when the distance-based modal split is considered. Consequently, the car mode receives a higher proportion (75.7%) as the car trips are on average longer.

It is interesting to have a look at the modal shares conditional on the trip distance. Modal shares categorized by six mutally exclusive distance bands are presented in Figure 2.7. The most significant difference between distributions can be noticed with-in the first trip length band (0–5 km) where the distance-based assessment assigns lower modal share to walk trips and consequently a higher share to car and public transport trips.

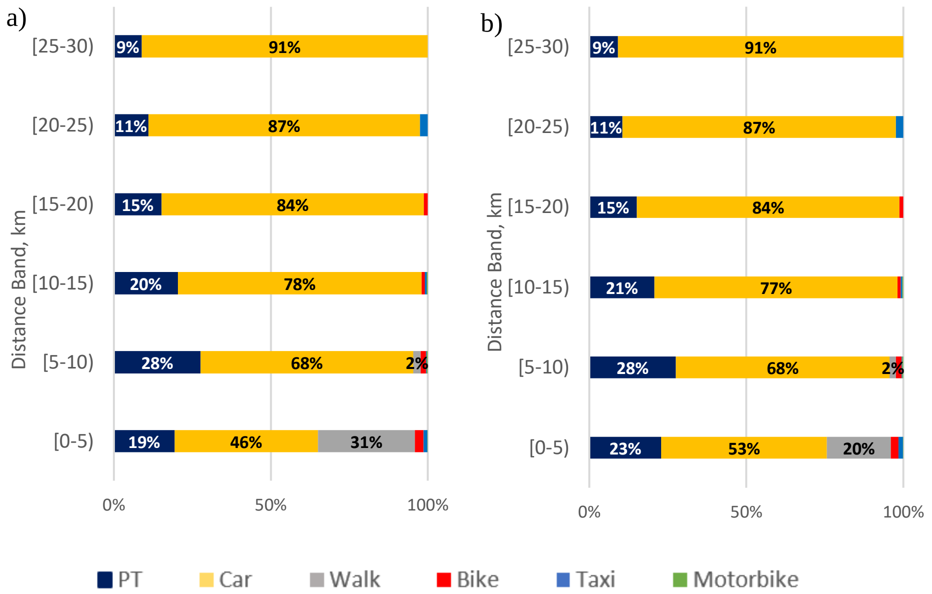

Fig. 2.7. Modal Split by Distance Bands: a) Trip-based; b) Distance-based (Created by Author)

Further, an investigation of MS conditional on the levels of the residents' characteristics is presented (see Table 2.5) and discussed.

Table 2.5. Modal Split (%) Categorized by Residents Characteristics

Varia-ble	Level	PT		Car		Walk		Bike		Taxi		Motorbike	
		T	D	T	D	T	D	T	D	T	D	T	D
Age	50 and over	29	30	53	66	17	3	1	1	0	0	0	0
	20–49	17	15	67	81	13	3	2	1	1	1	0	0
	16–19	31	37	37	48	23	7	5	3	1	2	3	3
Gender	Male	16	15	67	80	13	2	3	2	1	1	0	0
	Female	26	24	56	72	17	3	1	1	1	0	0	0
Education	Primary or Secondary Education	29	29	51	65	16	3	2	1	1	1	1	1
	Vocational Education	27	31	58	66	14	2	1	0	0	0	0	0
	Bachelor's Degree	14	11	69	84	13	3	3	2	1	1	0	0
	Higher Education	23	22	64	76	13	2	0	0	0	0	0	0

End of Table 2.5

Varia-ble	Level	PT		Car		Walk		Bike		Taxi		Motorbike	
		T	D	T	D	T	D	T	D	T	D	T	D
	Advanced Degree	21	18	60	77	16	3	2	2	1	0	0	0
Occupation	Unem-ployed	12	14	69	83	18	3	1	0	0	0	0	0
	Employed	18	17	65	79	14	3	2	1	1	0	0	0
	Retired	43	43	28	48	24	4	5	5	0	0	0	0
	Student	50	52	31	39	13	4	3	2	2	3	1	1

Abbreviations: T – Trip-based Modal Split, D – Distance-based Modal Split, PT – Public Transport

There are several trends to note:

a) young and elderly people tend to use public transport and walk more often than middle aged residents;
b) females choose public transport and walk more often than males;
c) the more educated the urban resident is, the more likely that a car is chosen for the trip making.

Two-factor analysis is given in Table 2.6, where the variation of modal split across groups of all possible age and occupation combinations is presented and discussed. Unemployed and employed people can be identified as the main car users. Among the employed residents, the ones between 20 and 49 seem to be reliant on car usage the most. Conversely, among the unemployed people those over 50 travel with the car extensively.

The above observations suggest that more sustainable modes (public transport, walk, and bike) are being used mainly due to the financial constraints and a car remains the dominant choice for those who are economically stronger and prefer a higher freedom of movement. However, such a strong reliance on the car mode results in frequent congestion and consequently environmental and social costs such as air and noise pollution, high energy consumption, road accidents, and delays. To move towards more sustainable mobility, some proactive measures are necessary. In general, sustainable mobility is a broad definition and according to Banister (2008) it encourages not only modal shift, but also the reduction of travel, and greater efficiency in the transport system.

The main approach that helps in seeking the sustainability of transportation systems is Travel Demand Management (also known as Mobility Management), which aims at promoting sustainable transportation by changing traveller behaviour (Grigonis *et al.* 2014; Santos *et al.* 2013; Kepaptsoglou *et al.* 2012).

Table 2.6. Modal Split (%) Categorized by Age and Occupation

Age	Occupation	PT		Car		Walk		Bike		Taxi		Motorbike	
		T	D	T	D	T	D	T	D	T	D	T	D
50 and over	Employed	25	28	59	69	16	3	0	0	0	0	0	0
50 and over	Unemployed	12	4	76	95	12	1	0	0	0	0	0	0
50 and over	Retired	43	43	28	48	24	4	5	5	0	0	0	0
16–19	Employed	2	2	58	82	29	9	4	2	0	0	6	5
16–19	Student	43	53	29	33	21	7	5	3	1	3	2	1
20–49	Employed	16	13	67	82	13	3	2	1	1	1	0	0
20–49	Unemployed	12	17	68	80	20	3	1	0	0	0	0	0
20–49	Student	61	50	35	48	1	0	0	0	3	2	0	0

Abbreviations: T – Trip-based Modal Split, D – Distance-based Modal Split, PT – Public Transport

It is a common knowledge that no single measure can make a difference and a set of push and pull measures should be used to relieve the situation. Habibian *et al.* (2011) identifies that pull policies encourage the use of non-car modes by making them attractive to car users; these policies include transit-oriented development, street reclaiming and the development of bus rapid transit. Inversely, push policies are those that discourage car usage by making it less attractive; these policies include road tolls, parking fees and cordon pricing.

A very well developed review of the full range of measures is compiled by Litman (2003) and is made available for transportation professionals, politicians and the general public via the online Travel Demand Management Encyclopedia (Victoria Transport Policy Institute 2003). This information has been reviewed by experts and is regularly expanded and updated.

Having in mind the geographical, social and political situation, Vilnius would benefit from further fine-tuning of the bus rapid transit coupled with transit-oriented development and restrictions on downtown parking. At this stage, restrictions on the use of polluting vehicles would also improve the situation.

2.4.6. Identification and Analysis of the Empricial Distribution of Trip Purposes

An analysis of people activities undertaken throughout the course of the day allows for a better understanding of the needs of contemporary urban residents.

During the survey residents were asked to identify the purpose of their trips or in other words, activities to be undertaken at the end of the trip. Throughout data cleaning and transformation, all the activities were assigned to one of the predefined categories identified in Table 2.7.

Table 2.7. Classification of Trip Purposes/Activities and their Abbreviations

No.	Activity Code	Activity	Activity Type
1	W	Work	Subsistence
2	C	Communication/Social	Discretionary
3	E	Educational	Subsistence
4	B	Business	Subsistence
5	S	Shopping	Maintenance
6	L	Leisure	Discretionary
7	H	Healthcare	Maintenance
8	A	Athletics	Maintenance
9	O	Outdoors	Discretionary
10	M	Maintenance Activity/Staying at Home	Maintenance
11	D	Drop Off (At School)	Subsistence
12	K	Drop Off (General)	Maintenance
13	P	Pick Up (School)	Subsistence
14	Z	Pick Up (General).	Maintenance
`15	F	Car Maintenance (Refuelling etc.)	Maintenance
16	X	Not Identified	-

In total, there are fifteen activities reflecting the most common endeavours undertaken by sampled respondents. The first ten categories identify temporary longer activities, which usually take place from between half an hour to several hours, with the next five activities being of a short term nature and taking place for up to just half an hour. There was a need to allow unknown activities (X) as some respondents have provided no information on the trip purpose. Other authors are using a significantly more aggregated classification. For example Reichman (1976) and Krizek (2003) classify the activities into three main categories:

a) subsistence activities that includes work, school or college related activities;

b) maintenance activities that includes personal, appointment, and shopping related activities;

c) discretionary activities that includes social visits and free-time.

Unfortunately, when considering only these three types of activities, lots of useful information is lost and the analysis suffers from less precise and sometimes even slightly obscure insights. Therefore, the full set of 16 activities will be maintained throughout this 2nd chapter and later in the modelling stage. It is worth noting that only up to 4 trip purposes are usually considered in conventional trip-based models, which consequently are inferior to the tour-based models taking into account daily trip sequences.

Figure 2.8 summarizes the empirical frequencies of activities (trip purposes) for the overall sample.

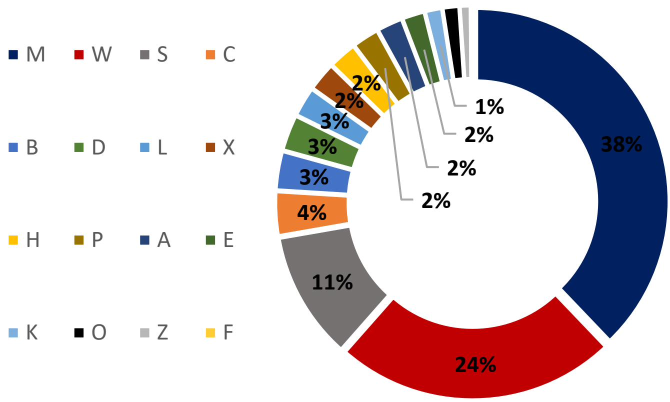

Fig. 2.8. Empirical Activity Frequencies (Created by Author)

By looking at the overall distribution, it is seen that the most likely purposes are related to Home (M), Work (W) and Shopping (S). These three purposes alone account for 69 % of all the trips. However, it is worth noting at this point that this distribution applies to the population older than 16 years and therefore there is a rather low percentage of Education (E) trips, which would have been significantly higher had the whole population been considered. Table 2.8 provides more detailed information on the distribution of activities categorized by residents' characteristics.

There are no very clear distribution differences in terms of gender. In general, females are slightly less likely to undertake work activities, more likely to undertake social communication and pick up (school related) activities and less likely to do business related trips.

Table 2.8. Empirical Activity Frequencies Categorised by Residents Characteristics

Variable	Level	Empirical Activity Frequencies, %															
		W	C	E	B	S	L	H	A	O	M	D	K	P	Z	F	X
Gender	M	26	3	2	5	10	2	2	2	1	38	2	2	1	1	0	3
	F	22	5	1	3	11	3	3	2	2	39	3	1	3	1	0	3
Education	AD	24	4	1	3	11	2	2	3	2	37	3	1	2	1	0	3
	BD	24	4	0	4	12	3	1	2	2	35	3	2	3	1	0	2
	HE	27	2	0	3	9	1	4	2	1	41	2	1	2	1	0	3
	PSE	19	4	7	3	10	2	4	2	1	42	1	1	1	0	0	2
	VE	26	2	1	3	8	3	3	0	1	41	4	2	2	2	0	2
Age	20–49	25	3	1	4	10	3	2	2	2	38	4	1	3	1	0	2
	> 50	22	4	0	3	13	2	4	2	1	40	1	1	1	1	0	4
	16–19	10	8	18	1	8	7	1	4	1	39	0	2	0	0	0	2
Occupation	U	4	4	2	1	12	3	5	1	6	42	9	1	6	0	0	2
	E	28	3	0	4	10	2	2	2	1	37	2	1	2	1	0	2
	R	1	6	0	0	21	2	10	4	2	44	1	1	1	1	0	7
	S	6	7	22	1	9	5	2	3	1	41	0	2	0	0	0	2

Abbreviations: VE – Vocational Education, PSE – Primary or Secondary Education, HE – Higher Education, BD – Bachelor Degree, AD – Advanced Degree, M – Male, F – Female, U – Unemployed, E – Employed, R – Retired, S – Student.

An examination of results related to education levels concludes with a rather trivial statement that less educated people do less work related and more education related trips with other trip purposes having relatively standard values.

The age characteristic allows us to distinguish different travel behaviours in terms of undertaken activities/trip purposes. Young people undertake a relatively higher proportion of social communication and education related trips and a lower proportion of work and shopping trips, which is a rather expected outcome. Furthermore, people aged 50 and over tend to do more healthcare related trips, more shopping trips and less work trips.

The occupation categorical variable also seems to influence the observed activity frequencies. Unemployed people show a relatively high proportion of shopping, outdoors, drop off and pick up activities. Retired individuals can be characterized by frequent shopping, healthcare, and social communication activities. Employed individuals feature a high frequency of work related and shopping trips whilst students conduct education, shopping and social communication related trips most of the time.

2.4.7. Identification and Analysis of Trip Lengths

Unfortunately, trip lengths have not been surveyed directly from respondents and that posed a significant issue during the analysis of trip distances. However, practical experience shows that even if distances had been observed, the reported estimates would have been significantly biased due to the respondents' inability to evaluate the travelled distance accurately.

However, the respondents were asked to identify the approximate addresses of their undertaken activities and this piece of information was used to estimate the distance between a properly defined origin and destination locations. Origin and destination locations were fed into the computational procedure that allowed the estimation of the shortest distance and travel time in the congested transport network between defined origins and destinations via Google Maps Distance Matrix Application Programming Interface (Google LLC 2018).

The procedure consists of the two main steps:

1. Geocoding of the addresses using Geocoder library (Carriere 2018) written in Python language. This library allowed the conversion of the addresses into a set of geographical coordinates: latitude and longitude in WGS84 coordinate system. It is worth noting that significant checking and correction efforts have been made to ensure that the geocoded locations were sensible.
2. Trip distance estimation. As soon as the locations have been identified properly the search of the shortest path was carried out with the help of Google Maps Distance Matrix Application Programming Interface following the procedure set out by Dumbliauskas *et al.* (2017) and Wang *et al.* (2011). Within this step, an explicit assumption that travellers have chosen the shortest route was made, which is not necessary true in all the observations. However, bearing in mind that this is the only way to estimate travelled distances, the procedure is deemed to be fit for purpose.

Figure 2.9 provides average trip length (ATL) estimates for various transport modes. The data reveals that the ATL of all the trips (4,049) in the sample is 7.5 kilometres. From a comparison of separate modes, car trips are about 30% longer than public transport trips and cycling trips are approximately two times shorter than car trips. It is worth noting that walking trips are comparatively long with an average distance of 2.0 kilometres.

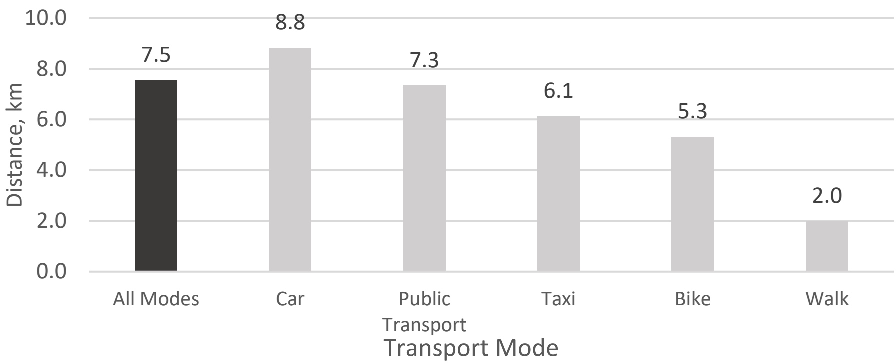

Fig. 2.9. Average Trip Lengths by Mode, km (Created by Author)

Table 2.9 provides a more detailed analysis of the ATL parameter categorized by respondents' characteristics and mode.

Table 2.9. Average Trip Lengths (km) Categorized by Residents' Characteristics and Mode

Variable	Level	All Modes	Public Transport	Car	Walk	Bike	Taxi
Age	50 and over	7.0	7.0	8.4	1.6	7.2	-
	20–49	7.8	7.5	8.9	2.1	4.9	5.4
	16–19	7.5	8.2	10.0	2.5	4.3	23.6
Gender	Male	7.9	7.3	9.2	1.9	4.8	7.0
	Female	7.2	7.4	8.4	2.1	7.3	5.1
Education	Primary or Secondary Education	8.4	8.1	10.3	2.0	5.4	-
	Vocational Education	8.4	8.7	9.9	1.7	-	-
	Higher Education	8.3	9.1	9.0	1.9	-	-
	Bachelor's Degree	7.1	6.8	8.1	2.2	5.0	4.9
	Advanced Degree	7.1	6.1	8.6	1.9	5.7	4.0
Occupation	Unemployed	8.2	8.6	9.4	1.6	-	-
	Employed	7.6	7.2	8.8	2.1	4.9	5
	Retired	6.2	7.5	7.9	1.3	7.7	-
	Student	7.8	7.6	10	2.5	4.8	-

By looking at the age levels it is obvious that across all modes the longest trips are being made by middle aged (20–49) individuals. Furthermore, it is seen that males (7.9 km) tend to take longer trips than females (7.2 km); however, the difference is rather small. Consideration of ATL across education levels brings a conclusion that more educated people make shorter trips. It has been noted previously that more educated people make more trips, so it seems that generally this group tries to maximize their number of activities while minimizing the time spent for travelling. Finally, it should be noted that the unemployed people make the longest trips whereas retired individuals make the shortest ones. The derived trip lengths allowed the identification of trip length distributions, which is the critical piece of information within the calibration of the trip distribution sub model. Modelled trip distances will be compared to the observed ones and trip distribution model parameters will be optimized to achieve a reasonable representation. The following graph in Figure 2.10 illustrates trip length distribution by time intervals within the whole sample, irrespective of the trip purpose.

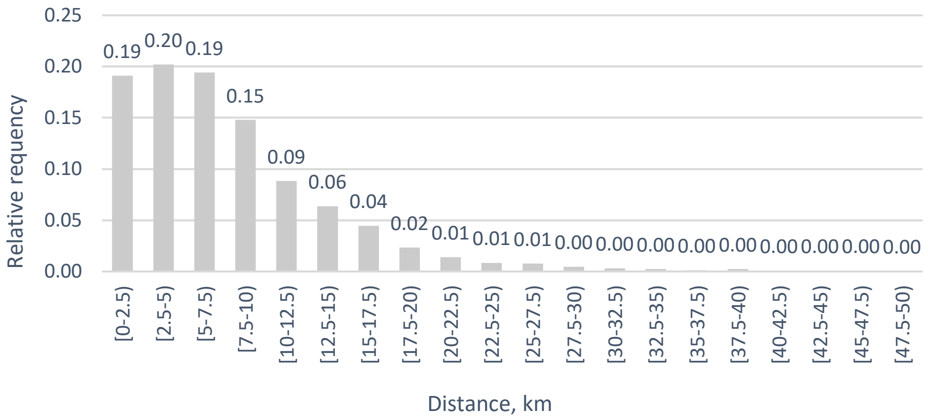

Fig. 2.10. Trip Length Distribution by Distance Intervals for All Activities/Trip Purposes (Created by Author)

The distribution reveals that about 60% of trips are not longer than 7.5 kilometres, about 80% of trips are not longer than 12.5 km and the average trip length is 7.7 km. This is rather typical shape of the trip length distribution that is also found across many towns and cities. It has an exponential shape significantly skewed to the right and its range usually depends on the size of the city or town. The three figures that follow (Figure 2.11 to Figure 2.13) define trip distributions for Work, Shopping and Home Maintenance activities.

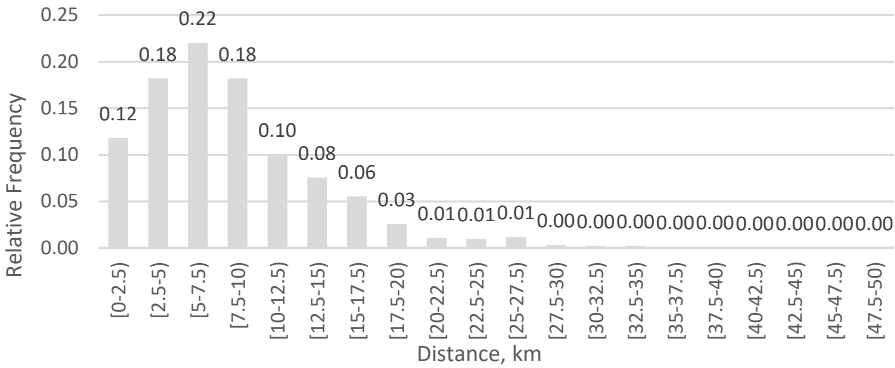

Fig. 2.11. Trip Length Distribution by Distance Intervals for Work Activity (Created by Author)

Researchers (Fellendorf *et al.* 2000) have found work related trips to be less sensitive to separation/distance compared to other trips. And this statement is backed up by the outputs of our survey. The average trip length is 8.4 km and it is far above the overall average (7.7 km). Compared to overall distribution it has more mass to the right. The distribution reveals that about 52% of trips are not longer than 7.5 km and about 80% of trips are not longer than 12.5 km.

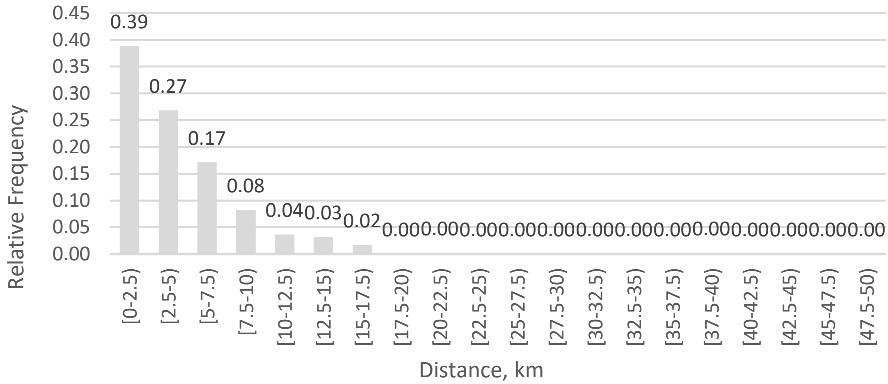

Fig. 2.12. Trip Length Distribution by Distance Intervals for Shopping Activity (Created by Author)

Distribution related to shopping activity features a significantly different shape and resembles exponential distribution. The average trip length is 4.4 km, which is far below the overall average. This difference is mainly due to the shopping activity being rather flexible and not fixed to a particular location over time (short term decision) as well as due to the availability of shopping centres spread all over the place.

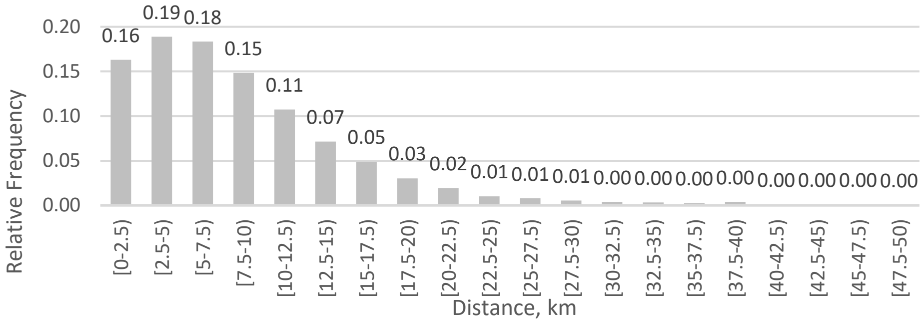

Fig. 2.13. Trip Length Distribution by Distance Intervals (km) for Home Maintenance Activity (Created by Author)

The average home related trip length is 8.4 km, which is slightly above the overall average and equal to the average associated to work related trips. About 54% of trips are not longer than 7.5 km and about 79% of trips are not longer than 12.5 km.

Trip length distributions can be approximated by Gamma probability density function, expression of which is given below:

$$f(x) = \frac{x^{\alpha-1} \cdot e^{-\frac{x}{\beta}}}{\beta^{\alpha} \cdot \Gamma(\alpha)}, \tag{2.6}$$

here α, β – distribution parameters; x – trip length; $\Gamma(\alpha)$ – gamma function.

The gamma function $\Gamma(\alpha)$ is an extension of the factorial function to real numbers and is calculated as follows:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} \cdot e^{-x} dx. \tag{2.7}$$

The graphs given in Figure 2.14 presents gamma distributions fitted using least squares methodology for work, shopping, home and all trips in combination.

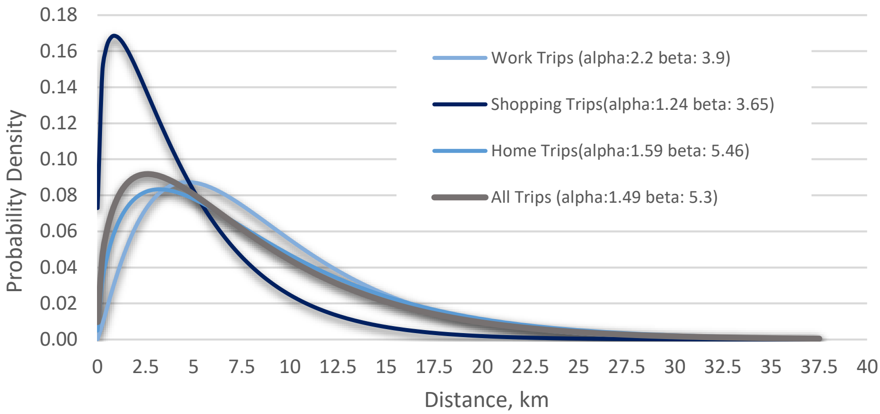

Fig. 2.14. Gamma Distributions Fitted to Trip Length Distributions (Created by Author)

The fitted distributions closely follow observed ones and here again a major difference between shopping and other trips can be identified.

2.4.8. Identification and Analysis of Trip Start Times

Within this section, the temporal distribution of trip starting times will be examined. First the distribution of the overall sample of trips is presented and then a more detailed analysis of the distributions by activity follows. The chart shown in Figure 2.15 describes the relative frequencies of trip starting times for each hour during the day.

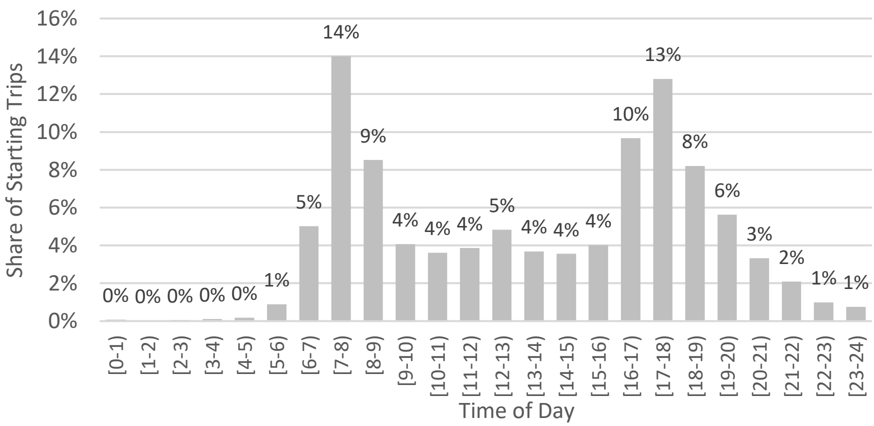

Fig. 2.15. Distribution of Trips by Start Time (Created by Author)

There are two clear global AM and PM peak periods and a rather flat profile of trips between them. The most intensive hour, during which about 14% of trips are made, belongs to the AM peak and is between 7:00 and 8:00 o'clock. The next most intensive hour containing about 13% of trips lies within PM peak and is between 17:00 and 18:00 o clock. It is worth noting that out of the five most intensive hours two belong to AM peak and three belong to PM peak. This suggests that even though AM peak has higher short-term intensity, PM peak takes place longer.

Further, the distribution of trips by start times and purposes is given in the Annex A. As can be seen from data, about 79% of work (W) related trips start between 6:00 and 9:00 with an AM peak hour containing 42%. Shopping (S) trips are made mainly during lunch time and immediately after work. It can be seen from the table below that the period between 11:00 and 13:00 contains about 19% and the period between 16:00 and 19:00 holds about 40% shopping (S) trips. About 60% of home related trips take place during the four hours (16:00–20:00) located within the PM peak period. School related drop-off (D) trips mainly (82%) are undertaken during two AM peak hours between 7:00 and 9:00, whereas pick-up (P) trips are slightly more dispersed within PM peak period with 82% of trips being located between 16:00 and 19:00. It is important to recognize that car maintenance related trips (F) and trips without any defined purpose (X) have had very low samples, therefore their frequency estimates are unreliable.

Temporal distributions of the trips are a key survey output as they will be directly used to model temporal distribution of activities within the tour-based travel demand model.

2.4.9. Identification and Analysis of Activity Sequences

This section describes the final piece of the analysis conducted within the scope of this empirical survey. Here, the observed activity sequences and their various statistics are presented.

Observed activity sequences and, their relative frequencies with associated probabilities for the overall sample are presented in Table 2.10. It is worth noting that the high number of different sequences were identified during the analysis and listing all of them in the following table was just not feasible. Therefore, the list presents only the 30 most likely different sequences. Together these form the basis of all the resident-related mobility taking place in Vilnius City, as all other sequences contribute towards mobility to a significantly smaller extent (321 observed sequences out of 1979 cases).

Probabilities were estimated by following a rather simple procedure: a number of observations of a specific activity sequence made by members of the sample is divided by the total sample size (n =1773 residents). The probability represents the likelihood of a typical member of the population conducting the given activity sequence throughout the course of a typical weekday.

Table 2.10. Observed Activity Sequences

Observed Activity Sequences	Number of Observations	Relative Frequency	Probability
Staying at Home	424	21.4%	0.239
MWM	525	26.5%	0.296
MSM	122	6.2%	0.069
MWSM	86	4.3%	0.049
MHM	48	2.4%	0.027
MCM	46	2.3%	0.026
MEM	42	2.1%	0.024
MLM	38	1.9%	0.021
MXM	37	1.9%	0.021
MOM	29	1.5%	0.016
MBM	26	1.3%	0.015
MWBM	26	1.3%	0.015
MAM	25	1.3%	0.014
MDM	25	1.3%	0.014
MDWPM	22	1.1%	0.012
MWCM	18	0.9%	0.010
MDWM	16	0.8%	0.009
MPM	12	0.6%	0.007
MWAM	11	0.6%	0.006
MWLM	10	0.5%	0.006
MWXM	10	0.5%	0.006
MKM	8	0.4%	0.005
MBSM	7	0.4%	0.004
MCSM	7	0.4%	0.004
MHSM	7	0.4%	0.004
MSSM	7	0.4%	0.004
MWASM	6	0.3%	0.003
MWBSM	6	0.3%	0.003
MWPM	6	0.3%	0.003
MWSCM	6	0.3%	0.003
All Other Sequences	321	16.2%	0.181
Total:	1979	100.0%	1.116

Table 2.11 contains one dummy activity sequence entitled “Staying at Home”, which is not an actual sequence but just a placeholder for a probability of not travelling, and is given just for clarity and convenience. As one individual can undertake more than one activity sequence during any given day, the probabilities sum to more than one. For example, a person can choose to travel to work, get back home and then do another tour by travelling to the local supermarket. Probabilities are a key to tour-based travel demand modelling as the multiplication of probabilities with the total population size allows the estimation of the total number of conducted sequences.

It is seen from the table that most of the sequences are single purpose and only few of them are complex combining multiple activities. There is no surprise that the two most frequent simple sequences relate to work and shopping activities as these are undertaken daily most of the time.

The eight most likely complex sequences (MWSM, MDWPM, MWBM, MWCM, MDWM, MWXM, MWLM, MWAM) involves work as one of the activities, which highlights that people tend to supplement their work-related travel with additional activities en route. De Abreu e Silva (2018) argues that chaining trips and having a smaller number of more complex tours during a day is considered as an individual strategy to reduce the total amount of travel, particularly total distances travelled.

There is an evidence that the complexity of an activity sequence is highly influenced by land use patterns and increases over time due to the changes in a lifestyle. In addition, there is a potential association between more complex travel behaviour and dependence on car use.

Ma *et al.* (2014) finds that complex tours are usually done by people living in a low-density, mono-functional environment located further from the central area. Similar findings has been documented by Krizek (2003) who concluded that households living in areas with higher levels of neighbourhood access are found to complete more tours but make fewer stops per tour.

Some authors (McGuckin *et al.* 2005; Levinson *et al.* 1995) have found in their longitudinal studies that activity sequences are becoming increasingly more and more complex over time. Activity sequences have increased in the past decades, in great part due to changes in the location of specific activities, which have moved from in-home to out-home (e.g., stopping for coffee or meals) and to escorting activities (mainly escorting children to school).

Ye *et al*. (2007) conjectures that complex activity sequences may lead to an increase in car usage. If one needed to pursue a complex sequence, then the flexibility afforded by the private automobile is desirable. The ability to pursue multiple activities in a single journey is rather limited when constrained by the schedules, routes, and uncertainty associated with public transportation.

Considering Table 2.11 given below, which identifies that about 69.2% of typical residents in Vilnius city do not undertake complex tours, it can be concluded that travel complexity is not the main reason for poor public transport usage. However, it should be noted that complexity has the potential to increase over time and the public transport system may face additional challenges in the near future.

Table 2.11. Distribution of Sampled Residents across Sequences and Activities

NDS	Number of Daily Activities, %									
	0	1	2	3	4	5	6	7	8	Total
0	23.9	–	–	–	–	–	–	–	–	24
1	–	40.3	14.2	7.2	2.5	0.7	0.2	0.2	0.1	65
2	–	–	4.6	2.2	1.7	0.8	0.3	0.3	0	10
3	–	–	–	0.3	0.2	0.2	0.1	0	0	1
4	–	–	–	–	0.1	0	0	0	0	0
Total	24	40	19	10	4	2	1	0	0	100

Abbreviations: NDS – Number of Daily Sequences

According to Table 2.11 above, the highest share of sampled residents pursues simple activity sequences (45.3%), e.g. people tend to conduct an activity and travel directly back home. The complex travel and activity sequences are featured by 30.8% of the sampled residents. The rest of the sample (23.9%) did not travel at all.

A presentation of sequences and their probabilities categorized by population segments is given in the Annex B. This table allows for the identification of key behavioural differences present between defined population segments. For example, the five most likely activity sequences associated with the retired person aged over 50 are: MSM (0.17), MHM (0.08), MCM (0.04), MXM (0.03) and MOM (0.01). In contrast, the employed person aged over 50 will most likely undertake the following sequences: MWM (0.42), MWSM (0.08), MSM (0.05), MXM (0.03) and MHM (0.02). The differences are apparent: retired persons have no incentive to undertake complex travel and their activities are concentrated on shopping, healthcare and social communication, wheres working individuals still do working and shopping activities most of the time and try to tie in compatible activies into one sequence. Various similar comparisons and analysis are possible based on the dataset given in the Annex B.

2.4.10. Discussion on the Data Collection Approach

Even though the travel behaviour data analysis had a focus on activities and their sequences rather than individual trips, the administration of the survey was carried out through the traditional web-based self report and face-to-face interview. It is worth noting that traditional methods (face-to-face interviews, telephone interviews, pencil-and-paper or web-based self reporting) have considerable drawbacks, such as (Shafique *et al.* 2015):

a) low accuracy due to dependence on memory and under-reporting;
b) low sample sizes due to high administration cost.

As it is difficult to remember minute details with certainty, accuracy of the collected data (departure and arrival moments, geographic locations, purposes) falls. Moreover, due to the same memory issue, small trips remain unreported most of the time.

Recently, an alternative data collection method (Bierlaire *et al.* 2013; Chen *et al.* 2015; Allström *et al.* 2017) employing smartphones has emerged. Since most smartphones are equipped with various sensors (GPS and accelerometer), and since smartphones are integrated in the daily life of most people, they provide an unprecedented opportunity for large-scale travel data collection. The method has numerous advantages:

a) very convenient for participants;
b) longer time span surveys;
c) high accuracy;
d) extensive datasets;
e) unbiased data;
f) less-costly due to high smartphone penetration.

This method is a viable alternative that can be applied in the national practice as smartphone penetration in Lithuania is over 70%. However, there are some potential challenges that requires a special consideration, such as recruitment of participants, and more importantly data privacy issues.

Recruitment of participants, for example, can be improved by suggesting small financial incentives (discounts on public transport trips, parking charges etc.) to members of the public that keep data collection applications running in the background on their smartphones. In addition, the proper data anonymization algorithms need to be employed in order to ensure positive public perception and compliance with General Data Protection Regulation. Such datasets are rather sensitive and should be collected, annonymized and used with exceptional care.

2.5. Conclusions of the Second Chapter

1. An innovative activity sequence-focused travel behavior research approach designed to collect data for tour-based travel demand model takes into account a set of 16 trip purposes, which is a significant im-

provement over conventional travel behaviour research approaches designed to cater trip-based models and typically considering 2–4 trip purposes.

2. Data collected under activity sequence-focused approach allows quantification of essential mobility parameters such as the proportion of travelling residents, the average number of trips, average trip lengths, mode splits as well as daily activity sequences and their probabilities.
3. After application of activity sequence-focused survey with a sample size of 1773, it was found that the Vilnius city resident conducts on average 2.3 trips on weekday with the average length of 7.5 km per trip. The chances of each trip being made by public transport, car or by foot on average are 21%, 61% and 15% respectively. Trips are related to work, shopping or home most of the time as these three activities have 24%, 11%, 38% average probabilities of being undertaken respectively.
4. The highest share of sampled residents pursues only simple activity sequences (45.3%), i.e. people conduct only sequences (one or several) that involve one activity and two trips: first from home to the activity's location and second from activity's location to home. In addition, 30.8% of the sampled residents conduct at least one complex sequence, whilst the rest of the sample (23.9%) did not travel at all.
5. The most likely simple sequence is the "Home–Work–Home" being made with a 29.6% average probability, whereas the most likely complex sequence is the "Home–Work–Shopping–Home" being done with an probability of 4.9%.
6. Empirical estimates of the above-mentioned variables and especially activity sequences and their probabilities will feed into the tour-based travel demand model, which is to be developed by employing PTV Visum macro modelling application.
7. Even though the travel behaviour data analysis had a focus on activities and their sequences rather than individual trips, the administration of the survey was carried out through the traditional web-based self report and face-to-face interview. This method has several disadvantages and more innovative approaches employing smartphone devices with built-in location and acceleration sensors will become more prominent in the near future.

3

Tour-based Model Development and Analysis of Model Application Results

This part of the thesis relies heavily on the results of previous chapters and describes the development and application the tour-based travel demand model. Theoretical findings from the 1st chapter facilitates in a choice of methods, whereas empirical findings from the 2nd chapter ensures that the model represents a relevant spatial, temporal and socioeconomic context.

In the first part of this 3rd chapter, data sources together with the tools used to manipulate them are presented. Spatial and temporal extent as well as resolution is considered, and this is followed by the discussion of model drawbacks and limitations that are important to understand when interpreting modelling results. Among the first things, the very important transport network development scheme is presented that will be implemented in the near future to alleviate current congestion levels and provide better accessibility.

A significant part of the 3rd chapter is devoted to the actual model development, where the demand and network modelling steps are described. Finally, the model is applied to test the impact of a new arterial street on travel demand and network performance.

Two scientific articles (Dumbliauskas *et al.* 2017a; Dumbliauskas *et al.* 2018) have been published on the topic of this chapter.

3.1. Analysis Tools and Data Sources

The following techniques and tools are employed in the development of the travel demand model:

1. MS Excel spreadsheets and Python scripting language for the preparation, management and analysis of numerical datasets.
2. QGIS geographic information system for the preparation management and analysis of spatial datasets.
3. PTV Visum travel demand modelling system for model development and calculations.

It is recognized that the selection of the tools has been mandatory and subject to personal taste. There are many viable alternatives that could have been used instead of the selected tools.

McNally *et al.* (2007) writes that in the field of transport research, nothing is more valuable yet simultaneously more limiting to the validation of theory and models than data. And indeed, travel demand modelling necessitates huge amounts of data describing socio-economic characteristics of transport analysis zones, mobility of individuals and spatio-temporal structure of transport networks.

The list of various publicly available and commercially distributed datasets employed in the model development is as follows:

a) population distribution and socioeconomic characteristics' spatial dataset maintained and distributed by The Lithuanian Department of Statistics;
b) travel diary survey dataset created in cooperation with Vilnius City Municipality Enterprise "Vilniaus Planas";
c) commercial entities spatial dataset maintained and distributed by State Enterprise Centre of Registers;
d) private transport network spatial dataset available from Open Street Map (OSM) website (Open Street Map Contributors);
e) public transport network spatial and temporal dataset available in General Transit Feed Specification (GTFS) format from Vilnius City Municipality Enterprise Communication Services (2018).

These datasets are inherently spatial and most of the time prepared, distributed, analysed and visualised using GIS tools and methodologies. This generally proves that GIS and transport research have always been and increasingly will be interrelated disciplines (Loidl *et al.* 2016). The abovementioned datasets are explained in more detail in the sections that follow.

3.2. Spatio-temporal Scope and Resolution

The model area (see Figure 3.3 and Figure 3.5 for the overall spatial scope and resolution) entails the territory of 3,704 square kilometres, 403 of which belong to the core Vilnius City Municipality administrative area. This core area is named as urban and is modelled with a more precise detail, whereas the territory outside the Vilnius City Municipality administrative area is named as suburban and is taken into account with a sole purpose to reflect some part of external trips within the urban area. The urban areas are comprised of 218 transport analysis zones whereas the suburban areas of 128 transport analysis zones. The exact spatial extent of the urban and suburban areas with the delineated transport analysis zones is given in Figure 3.3 and Figure 3.5 in the following chapters. As can be deduced from these figures, suburban transport analysis zones encompass some outer territories that have close ties with Vilnius, such as: Salcininkai, Elektrenai, Trakai and Sirvintos Districts.

As far as time is concerned, the model encompasses a typical weekday and is based on a travel survey that has estimated whole day travel patterns. In addition, a one-hour resolution is used to model fluctuations within travel demand and supply. This essentially means that modelling results (volumes, travel times, vehicle kilometres travelled) can be presented not only aggregated for the whole day, but also for hourly time slices.

3.3. Initial Assumptions

Due to the lack of available data, some simplifying assumptions had to be made which put some limitations on the model's ability to reflect reality and in turn to predict future operation. These assumptions are explicitly recognized in the following list and need to be taken when interpreting modelling results:

1. Trips undertaken by population members under the age of 16 are not considered. During the empirical travel behaviour survey, which is described in 2nd chapter, the data regarding people younger than 16 years was not collected due to the legal constraints, and thus the modelling of this group's travel is not undertaken. A potential compromise around this issue would be to assume that the behaviour between this group and the group of age of 16 to 19 – for which the data has been collected – is not

significantly different. However, there is no reasonable evidence to back up this statement, so this population group is excluded from the analysis entirely. This assumption will cause a slight underestimation of trips as this group comprise about 15% of population.

2. Travel behaviour in urban and suburban areas is assumed to be identical. Here the term urban refers to all transport analysis zones located in Vilnius Municipality administrative area and the term suburban to all other zones. This assumption is needed due to the lack of data about travel behaviour of suburban residents, as the survey was administered only to the Vilnius municipality area. It is worth recognising that the main purpose of this project is to reasonably model urban zones and all suburban zones present in the model are considered with a sole purpose to ensure a reasonable number of external trips within the urban zones.
3. Trips with the endpoints outside the urban and suburban area are not considered at all. No roadside interview data is available at the edges of the suburban area and therefore these trips cannot be reflected reasonably within the model.

3.4. Scenario under Assessment

There are various natural phenomena and human implemented policies/actions influencing travel demand and network operation. Some of them can be directly addressed within this travel demand model:

1. Shifts in demographics: new residential developments, an aging society etc. As more and more new house developments are being built in suburban areas, the young families are settling down on the outskirts of the city. This phenomenon can be directly assessed within this travel demand model by manipulation of population levels within the suburban transport analysis zones. As a result, the number of tours starting and ending at these zones are going to increase by creating additional demand for travel. In a similar fashion, an aging society technically means a higher percentage share of the population within older age classes within each (or some) transport analysis zones and this effect will reduce the demand for travel as the observed mobility of elders has been found to be less intense compared to the younger population. The model is capable of estimating the impact on travel demand and network performance; however, the quantification of the demographic changes should come from external forecasts.

2. Spatial displacement, addition or deletion of activity locations: new shopping centres, industrial sites, large office buildings etc. New developments create opportunities for people to come and undertake various activities such as work, shop, communicate or relax. The travel demand model allows for predicting their impact on the travel demand redistribution and network performance; however, the quantification of the available new opportunities for activities should come from external forecasts.
3. Development of the transport network/supply: new transport systems, increased quality, improved connections etc. Improvement to transport system has a twofold effect: first some of the transport analysis zones become much more attractive as their accessibility improves; second, the routes traversing new infrastructure also get more attractive due to the lower travel times, higher comfort levels etc. The model can assess the impact of the network changes on the travel demand and network itself; however, as before, the changes to supply should come from the external sources such as a master plan or similar.

For demonstration purposes the tour-based demand model is employed in the assessment of the impact of the new Siaurine Street, which is due to be built in the northern part of Vilnius city. This case corresponds to the third bullet point in the above list and it was chosen for a number of reasons:

1. The high degree of certainty. The first idea of this link was presented several decades ago and now the decision taken of building it is well established within the Master Plan and political environment. As a result of this, the testing of the street within the demand model is well timed and allows a better understanding of what happens with demand patterns and network performance when the works are finished.
2. The extent of intervention. The street will connect several main arterials such as the Western Bypass, Justiniskiu Street, Laisves Avenue, Ukmerges Street and Fabijoniskiu Street. In addition, the overall length (excluding slip roads) is said to be about 3.28 km and implementation will cost approximately 90 million euros. These facts are enough to realise the potential importance of the link on the demand patterns and network performance.
3. The need for quantification of project profitability. As the project entails a significant investment (approx. 90 million euros), cost-benefit analysis should be undertaken to ensure that the investment is worthwhile doing. Travel demand modelling allows quantification of travel time and vehicle kilometer savings that in turn could inform the cost-benefit analysis.

Figure 3.1 provides a schematic layout of the proposed infrastructure improvements.

Fig. 3.1. Schematic Layout of Siaurines Street (Grigonis 2018)

According to the layouts given in the the feasibility study (Grigonis 2018), the first implementation stage is going to connect the Western Bypass, Justiniskiu Street, Laisves Avenue, Ukmerges Street and Fabijoniskiu Street. The overall length will be 3.28 km and its cross-section will contain six lanes in total with four lanes dedicated for general traffic and two lanes for public transport vehicles. This arterial street will play a major role in serving east-west movements and most likely will increase the utilization of the Western Bypass.

Such a significant capacity improvement is also bound to have significant effect on the demand distribution and network performance. But one has to be cautious about the delivered benefits, because there is an ongoing debate in the literature as to how new transport arteries impact the performance of the transport system, environment and society. Blumenstock *et al.* (2017) finds that a new major connection diverts traffic from other more congested alternative roads and in addition increases speed of travel. However, this study measured the immediate effects resulting within one month of the road opening and care should be taken when expanding such a conclusion to the long-term period. For example, Duranton *et al.* (2011) find that congestion in U.S. cities is not reduced by new roads, because new roads in the long-run attract additional drivers, who then clog up the system. This finding is consistent with the economic theory, which suggests that over time, the increased convenience of travel will induce more travel, eventually leading to steady-state levels of congestion similar to the prior equilibrium.

To generalize the above findings, it can be stated that in the very immediate period (within several months of opening) new roads divert demand from other routes, modes or even departure times. And this provides significant congestion relief (vehicle kilometre and travel time savings) that generate considerable financial return. In the short-run, though, it can be expected that new infrastructure is going to generate additional kilometres travelled, because of induced demand, i.e.

an increase in travel resulting from improved travel conditions. This phenomenon is clearly unravelled by Gorham (2009) who clarifies that due to the reduced travel time people are expected to choose, for example, further away destinations. The principal scheme of the effect is given in Figure 3.2.

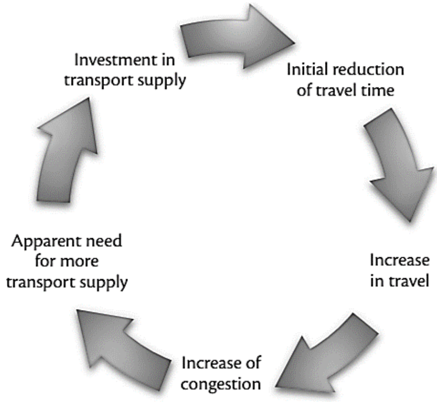

Fig. 3.2. Induced Travel's Vicious Circle (Gorham 2009)

Vehicle kilometres travelled can be expected to increase and congestion is likely to be nearing that of initial levels, and car users are expected to end up experiencing the same old congestion levels.

In the long run, increased adjacent land accessibility attracts new firms (addition of new activities) and companies and residential developments (demographic shifts) which in turn will generate new demand that can be thought of as a natural demand growth. This will benefit overall economic development as well as significantly increase congestion in the long run.

The tour-based demand model is capable of estimating the immediate and short-term effects, i.e. the diversion of travel from more congested routes and less attractive modes is accounted for as well as relocation of activities (potentially to further destinations) due to the reduction in travel times. However, the total amount of activity sequences remains constant, i.e. natural growth is not considered because of necessity of additional data (demographic shifting and appearance of new firms) that is not readily available.

In the following parts of this thesis, two scenarios are compared: Base Scenario without Siaurine Street and Network Development Scenario with Siaurine Street. The model is used to estimate the impact on network performance (vehicle kilometres travelled, travel time savings) and to predict the expected flows on the new infrastructure elements.

3.5. Travel Demand Estimation

3.5.1. Population Categories

Based on the travel behaviour analysis (see 2nd chapter), intuitive reasoning and data availability, the decision was made to segment the whole population by age and main occupation. As a result, the whole population is divided into eight segments, and it will further be assumed that the travel behaviour within these segments is homogeneous. To accomplish further travel demand calculations, the population distribution by each segment within each transport analysis zone is necessary. The data has been obtained from Lithuanian Department of Statistics (the most recent 2011 census dataset) in a spatial GIS format that is suitable for direct data transfer to transport modelling software PTV Visum. For information purposes, the aggregate population sizes by segment are given in Table 3.1.

Table 3.1. Population Segments and Their Sizes (Based on 2011 Census Dataset)

Age	Occupation	Urban Area		Suburban Area	
		Absolute Number	Relative Share	Absolute Number	Relative Share
16–19	Student	7,942	2%	2,148	2%
16–19	Employed	20,421	4%	5,524	5%
20–49	Employed	234,226	47%	51,239	45%
20–49	Unemployed	31,119	6%	6,808	6%
20–49	Student	12,503	3%	2,735	2%
Over 50	Employed	105,124	21%	25,377	22%
Over 50	Unemployed	9,574	2%	2,311	2%
Over 50	Retired	73,023	15%	17,628	15%
Total:		493,932	100%	113,770	100%

It can be seen from the table above that the three most prevalent groups are 20–49 years old employed, over 50 employed and over 50 retired. The size of these segments is very much influenced by the range of the age intervals. The spatial distribution of the overall population across transport analysis zones is given in Figure 3.3. It expresses the residential density as this measure is more meaningful in a spatial context. It can be noted that the density in the suburban transport analysis zones ranges from 10 to 900 with the average value of 150 residents per square kilometre.

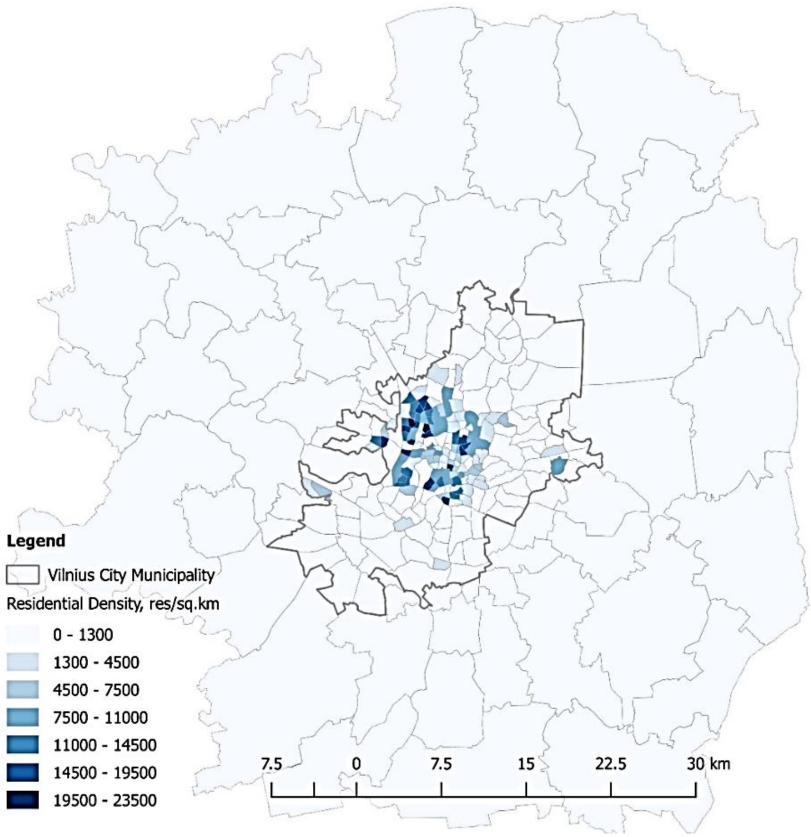

Fig. 3.3. Population Density Distribution across the Transport Analysis Zones (Created by Author)

In contrast, the density in urban zones varies between 0 and 16,000 with the average value of 3,800 residents per square kilometre. The most populous urban districts are Šeškinė, Žirmūnai, Justiniškės and Senamiestis (The Old Town).

3.5.2. Generation of Activity Sequences

The generation of activity sequences is the first calculation step of the travel demand model and it defines the demand for mobility. The population segment specific activity sequences and their probabilities (see 2nd chapter) together with population data are the key inputs within this procedure. Table 3.2 lists the five most likely activity sequences and their probabilities for each population segment. It provides some intuition about the activity sequences prevailing within the population.

Table 3.2. The Most Likely Activity Sequences for each Population Segment

Population Segment	Most Likely Activity Sequences and Their Probabilities				
	1	2	3	4	5
Over 50 and Unemployed	MSM (0.12)	MHM (0.06)	MWM (0.06)	MDM (0.03)	MOM (0.03)
Over 50 and Retired	MSM (0.17)	MHM (0.08)	MCM (0.04)	MXM (0.03)	MOM (0.01)
Over 50 and Employed	MWM (0.42)	MWSM (0.08)	MSM (0.05)	MXM (0.03)	MHM (0.02)
20–49 and Unemployed	MDM (0.14)	MSM (0.12)	MHM (0.07)	MPM (0.06)	MOM (0.05)
20–49 and Employed	MWM (0.40)	MWSM (0.06)	MSM (0.04)	MLM (0.02)	MCM (0.02)
20–49 and Student	MEM (0.29)	MWM (0.07)	MWSM (0.07)	MSM (0.02)	MLM (0.02)
16–19 and Employed	MWM (0.19)	MSM (0.14)	MWLM (0.10)	MEM (0.05)	MWSM (0.05)
16–19 and Student	MEM (0.39)	MCM (0.06)	MWM (0.04)	MLM (0.04)	MAM (0.04)

Technically speaking, multiplication of the sequence's probability with the total number of residents in the population segment allows us to obtain the total number of activity sequences. Calculations are carried out for each population segment within each transport analysis zone.

To avoid technical errors, shares of observed activity frequencies are compared to modelled activity shares. This check constitutes a simple quality assessment of the activity sequence generation procedure. The results are shown in the Table 3.3.

As all observed activity sequences (including the long and complicated ones) were coded within the model, the match between observed shares and modelled shares is almost exact. There have been no simplifications or omissions, and the minor evident discrepancies have been brought about by rounding errors within the data management and calculation procedures. The correctness of the results within this model step is highly related to the quality of sampling procedure. If the collected sample is representative of the whole population, then the results within this procedure are also valid.

Table 3.3. Comparison of the Modelled Activity Shares against the Observed Activity Shares

Code	Activity	Observed Shares	Number of Sampled Activities	Modelled Shares	Modelled Number
A	Athletics	2.3%	92	2.4%	32898
B	Business	3.5%	142	3.5%	47773
C	Communication	3.7%	151	3.7%	50744
D	Drop Off (At School)	2.7%	110	2.6%	35925
E	Education	1.5%	62	1.4%	18844
F	Car Maintenance	0.15%	6	0.15%	2080
H	Healthcare	2.4%	96	2.4%	32635
K	Drop Off (General)	1.4%	58	1.4%	18926
L	Leisure	2.5%	100	2.5%	34339
M	Maintenance	38.4%	1554	38.4%	530267
O	Outdoor	1.5%	62	1.5%	20344
P	Pick Up (At School)	2.0%	82	2.0%	26913
S	Shopping	10.6%	430	10.8%	148521
W	Work	23.7%	960	24.0%	330850
X	Not Identified	2.6%	107	2.7%	36834
Z	Pick Up (General)	0.9%	36	0.9%	12060

The data within the above table allows the quantification of the total number of activities (and consequently trips). The modelled population of people aged over 16 undertakes about 1.38 million activities (trips) over the course of the typical weekday. Work and shopping related activities constitute 24.0% and 10.8%, which amounts to some 331,000 and 149,000 activities respectively. It is once again worth noting that education activity is not properly represented within the model as the sample has purposefully excluded population members younger than 16 years old.

3.5.3. Distribution of Activities throughout the Day

It has already been mentioned in the 1st chapter that PTV Visum software models the distribution of activities over the time of day in a quite simplistic manner by using observed empirical distribution. More precisely, each activity is allocated a probability distribution over one-hour intervals within the modelled 24-hour period. The probabilities define the proportions of movements (activities) that are undertaken over the intervals within the modelled periods. As a result, the tour-based model reflects directly what has been observed in the sample and the discrepancies appear only due to the rounding errors in the data transfer and management procedures. Figure 3.4 provides the comparison of the observed and modelled distributions over the time of the day.

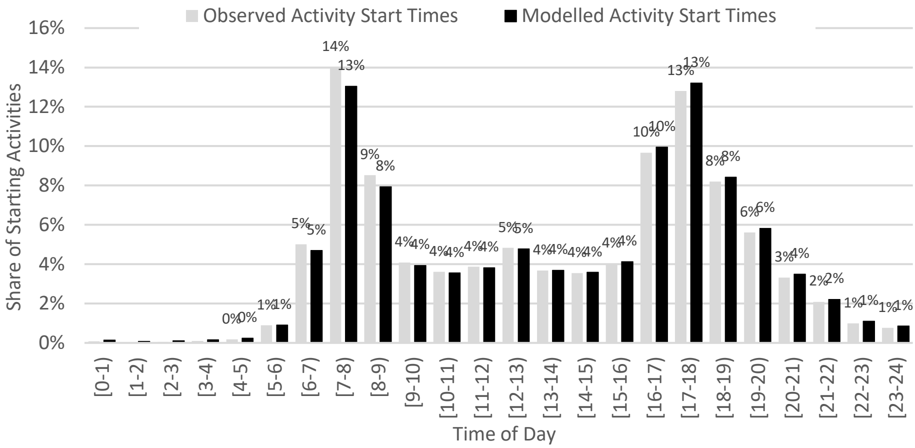

Fig. 3.4. Comparison of the Modelled Activity Start Times against the Observed Activity Start Times (Created by Author)

Once the hourly demand matrices are generated, the actual departure can take place either earlier or later compared to the desired departure, depending on the utility of the specific connections within the public transport or private transport dynamic assignment models.

3.5.4. Generation of Tours/Destination Choice

The purpose of this second calculation step is to assign spatial location (transport analysis zone) to each activity within the activity sequence. This generally means converting activity sequences that do not yet have any spatial dimension to location sequences. The key inputs to this modelling step are:

a) the outputs from the previous step: total number of activity sequences;
b) zonal attractiveness (structural property) values for each activity within each transport analysis zone;
c) spatial separation values between all zone pairs.

Zonal attractiveness expresses the potential of undertaking the particular activity within competing transport analysis zones. The dataset of the distribution of commercial entities across the spatial scope of the modelled area is employed for this task.

The commercial entities dataset is maintained and distributed by the National Centre of Registers. It contains the various attributes related to each entity, of which the address, the number of employees and the business area are the most important ones for the tour generation step.

The density of employment categorized by the commercial activity in each transport analysis zone is calculated and further used as a measure of attractiveness. Table 3.4 identifies the categories of employment used for each activity in the destination choice model.

Table 3.4. Mapping between the Activities and Zonal Attractiveness Variables

No.	Code	Name	Zonal Attractiveness
1	W	Work	Density of the overall employment
2	C	Communi-cation	Density of the employment in food services
3	E	Education	Density of the employment at universities
4	B	Business	Density of the overall employment
5	S	Shopping	Density of the employment in retail services
6	L	Leisure	Density of the employment in entertainment services
7	H	Healthcare	Density of the employment in healthcare services
8	A	Athletics	Density of the employment in athletics services
9	O	Outdoors	The same attractiveness is assumed for all zones
10	M	Maintenance	Not relevant. Trips are directed back to original zone
11	D	Drop Off (At School)	Density of the employment in preschool and elementary school services
12	K	Drop Off (General)	Density of the overall employment
13	P	Pick Up School	Density of the employment in preschool and elementary school services
14	Z	Pick Up (General)	Density of the overall employment
15	F	Car Maintenance	Density of the total number of fuel stations
16	X	Not Identified	Density of the overall employment

The overall employment density distribution across the transport analysis zones is given in Figure 3.5. The figure provides employment density as this measure is more meaningful in a spatial context and is used directly within demand model.

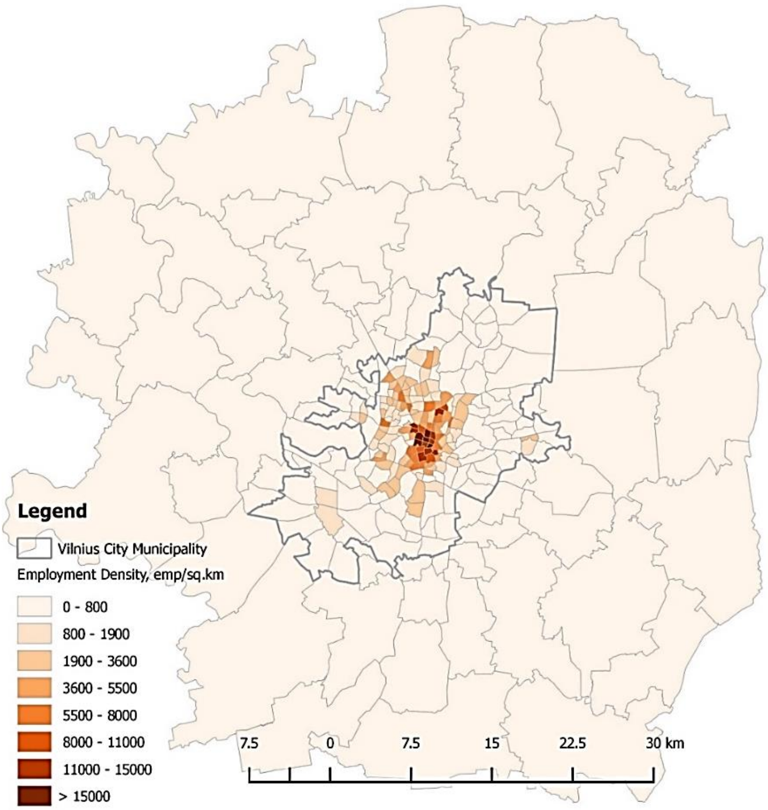

Fig. 3.5. Employment Density Distribution across the Transport Analysis Zones (Created by Author)

It can be noted that employment density in the suburban transport analysis zones ranges from 0 to 340 with the average value of 55 workplaces per square kilometre. In contrast, the density in urban zones varies between 0 and 44,000 with the average value of 3,000 workplaces per square kilometre. The most intensive areas are within the geometrical centre of the area of interest.

Spatial separation between the origin and destination pairs flows into the destination choice model. And in this demand model, spatial separation is expressed by the logarithmic sum (a.k.a. logsum) of the lower level mode choice utilities multiplied by the activity specific calibration coefficient β, which is varied to match the average trip lengths. The overall formulation of destination choice model is given below. Conditionally on the current location i, the probability that activity k, which is the next one in the activity sequence, will be undertaken at the location j is equal to the following:

$$P_{ij}^{k} = \frac{S_j^k \cdot e^{\beta^k R_{ij}}}{\sum_{m=1}^{B} S_j^k \cdot e^{\beta^k R_{ij}}}; \tag{3.1}$$

$$R_{ij} = ln \sum_{t=1}^{T} e^{IMP_{ijt}}, \tag{3.2}$$

here P_{ij}^k – probability that the next activity k will be undertaken at the zone j, conditional on the current zone i; S_j^k – attractiveness of activity k at the zone j; R_{ij} – separation between current zone i and potential destination zone j; IMP_{ijt} – impedance between current zone i and potential destination zone j by mode t; β^k – calibration coefficient; B – the set of potential destination zones; T – the set of available modes between current zone i and potential destination zone j.

The calculation of impedance between current zone i and potential destination zone j by all available modes ($IMP_{ijt}, \forall\, t$)will be explained in more detail in the section that follows.

A significant advantage of this tour-based destination choice model over the conventional trip-based model is that all trips within the tour are related. That is, the endpoint of the first trip within a tour becomes the origin to the second trip and so on and so forth.

The calibration exercise of a destination choice model seeks to reproduce the average trip lengths by each activity. This is achieved by varying the parameters β^k, which influence the propensity of travel to the distance. Higher coefficient values mean that travel is more sensitive to spatial separation and resulting trip lengths will be shorter.

To verify calculation results, modelled trip lengths are compared against the observed ones. Table 3.5 provides relevant information.

Table 3.5. Destination Choice Calibration Results and Quality Assessment

Activity	Modelled Average Trip Lengths, km	Observed Average Trip Lengths, km	Relative Difference
W	8.3	8.40	–1%
C	7.8	8.00	–3%
E	9.5	9.60	–1%
B	7.3	7.40	–1%
S	4.5	4.40	2%
L	9.4	9.70	–3%
H	7.2	7.10	1%
A	5.8	5.70	2%

End of Table 3.5

Activity	Modelled Average Trip Lengths, km	Observed Average Trip Lengths, km	Relative Difference
O	3.5	3.50	0%
M	8.7	8.40	4%
D	7.5	7.30	3%
K	7.6	7.60	0%
P	6.9	7.00	–1%
Z	6.5	6.80	–4%
F	7.6	7.5*	1%
X	6.9	7.10	–3%
W	8.3	8.40	–1%

* – overall trip length average

Most of the trip purposes show a satisfactory match with discrepancies lower than 5% between the modelled and observed values. It is worth noting here, that trips related to car maintenance activity (F) are compared to the overall average (7.5 km) instead of observed activity specific estimate. The issue with this set of trips is related to low sample size (only 6 trips recorded), which causes low reliability of the trip length estimate (13.6 km), and consequently, it can be far away from the true value. In addition, the model spatial extent does not allow for such a large average trip length, these car maintenance trips are calibrated to the overall car-based average, which is 7.5 km.

3.5.5. Mode Choice Step

The purpose of this step is to assign a mode to each trip between the activities. The mode choice step involves the Multinomial Logit model predicting the shares of the demand using public transport, car or taking a walk. The key inputs within this step are:

a) location sequences (sequences of transport analysis zones) as an output from destination choice model;
b) impedance values by mode as an artefact of the transport supply.

The mode within the location sequence can vary, but only across the set of modes that are defined as exchangeable (public transport and walk in this case). If the first trip within the location sequence is assigned to the non-exchangeable mode (car), no mode choice is conducted during the further trips within a sequence and car is assumed to be used over the whole tour.

The impedance is defined for each mode separately and is a combination of travel distance, travel time and alternative specific constants. The meaning of the time and distance variables varies by mode. For a car mode the distance measures the average length of all available paths between origin and destination pair and travel time measures the average trip duration over all the paths available for the origin destination pair in the loaded network.

Public transport trip distance and time components are the sums of the following terms that are averaged over all available connections between origin and destination pairs:

a) access distance/time that represents the walking from origin zone to the departure stop point;
b) in-vehicle distance/time that represents travelling with public transport vehicles;
c) transfer distance/time that represents the walking from one stop point to the next one during the transfer;
d) egress distance/time that represents walking from arrival stop point to the destination zone.

The model does not explicitly model the non-motorized network and therefore the estimation of walk distances and times is purely approximate. Walk distances between origin and destination pairs are represented with straight-line (Euclidean) distance between the centroids of the zones. In addition, walk times are estimated assuming the average walk speed of 4 km/h.

The impedance equations for each mode are given below. The origin and destination notation (i and j) is omitted for simplicity reasons.

$$IMP_{CAR} = \theta_{CAR} \cdot TT_{CAR} + \alpha_{CAR} \cdot \ln\left(\frac{D_{CAR}}{AD_{CAR}}\right) + ASC_{CAR}; \tag{3.3}$$

$$IMP_{PT} = \theta_{PT} \cdot TT_{PT} + \alpha_{PT} \cdot \ln\left(\frac{D_{PT}}{AD_{PT}}\right) + ASC_{PT}; \tag{3.4}$$

$$IMP_{WALK} = \theta_{WALK} \cdot TT_{WALK} + \alpha_{WALK} \cdot \ln\left(\frac{D_{WALK}}{AD_{WALK}}\right), \tag{3.5}$$

here $TT_{CAR}, TT_{PT}, TT_{WALK}$ – travel time by car, public transport and walk modes; $D_{CAR}, D_{PT}, D_{WALK}$ – distance by car, public transport and walk modes; $AD_{CAR}, AD_{PT}, AD_{WALK}$ – advantage distance by car, public transport and walk modes; $\theta_{CAR}, \theta_{PT}, \theta_{WALK}$ – travel time marginal effect on impedance by all modes; $\alpha_{CAR}, \alpha_{PT}, \alpha_{WALK}$ – logarithmic relative distance marginal effect on impedance by car, public transport and walk modes; ASC_{CAR} – alternative specific constant of car and public transport.

The parameter AD is called “advantage-distance“, because it specifies how long a trip must be so that distance has a positive impact on the impedance. It is intuitive that cars are more suitable for longer trips and walk and public transport modes usually are preferred for short trips over a car.

Travel time marginal effect coefficients θ, logarithmic relative distance marginal effect coefficients α, advantage distance parameters AD and alternative specific constants are varied to obtain the satisfactory match between observed mode shares and modelled mode shares.

To verify calculation results, modelled mode shares are compared against the observed ones. Overall observed modal shares are: 21.0% of public transport trips, 61.3% of car trips, 14.7% of walk trips and 3.0% of all other trips. In comparison, the tour-based model estimates 274,900 public transport trips, 888,800 car trips and 216,200 walk trips, which results into the following modal shares: 19.9% of public transport trips, 64.4%of car trips and the remaining 15.7% of walk trips as only three modes have been considered within a model. In conclusion, the aggregate match between the observed and modelled modal shares can be considered as very good.

In addition to above verification, collected data and associated analysis outlined in the 2nd chapter allows for a more precise comparison of observed and modelled mode shares within the five-kilometre distance bands. Table 3.6 provides relevant information.

Table 3.6. Mode Choice Calibration Results and Quality Assessment

Distance Band, km	Modelled Mode Shares			Observed Mode Shares		
	Public Transport	Car	Walk	Public Transport	Car	Walk
0–5	16%	54%	30%	19%	46%	31%
5–10	35%	65%	0%	28%	68%	2%
10–15	20%	80%	0%	20%	78%	0%
15–20	17%	83%	0%	15%	84%	0%
20–25	6%	94%	0%	11%	87%	0%
25–30	1%	99%	0%	9%	91%	0%

As can be seen from the table the match between the observed and modelled values within the 5 km distance bands can be considered as good. with discrepancies being up to 8% within any single cell. These errors occur due to a number of reasons such as imprecise model specification, sampling variation etc. Bearing in mind the overall mode share match, the errors in the distance bands can be neglected and the model is considered as fit for its purpose.

3.6. Network Model Development

Transport network development is a rather tedious and time-consuming task, especially when the network requires the need to be coded accurately for a large spatial extent. Recently many open data sources have emerged and become available for use in transport models. The appearance of these open data structures has been either driven by voluntary efforts (for example OSM) and the opportunities of added value created to public agencies by third party developers. On the other hand, one should not neglect the positive impact of recent amendments to the Public Sector Information (PSI) Directive (European Parliament 2013) that supported and promoted the wide availability and re-use of public sector information for private or commercial purposes, with minimal or no legal technical or financial constraints. Such legal requirements create a groundwork for how open data should be incorporated into the culture of public sector.

Within the context of this thesis, the two most relevant open data sources and Open Street Map (OSM) database and General Transit Feed Specification (GTFS) database. The first dataset was used to create the Visum private network structure, whereas the second one was transformed into the Visum public transport network model. The datasets and procedures are defined in more detail in the following subsections.

3.6.1. Private Transport Network Development

Recently, Web 2.0 technologies and the wide dissemination of GPS-enabled devices boosted public participation in collaborative mapping projects and the OpenStreetMap (OSM) project has been one of the most successful representatives (Arsanjani *et al.* 2015). This type of collaboratively collected spatial information was termed Volunteered Geographic Information or in short VGI by Goodchild (2007).

OSM data is free and available for personal, group and commercial use under the Open Data Commons License, which provides for users to copy, distribute, modify and build upon a database if the proper attribution is in place (OpenStreetMap Wiki Contributors 2016).

Since the outset of the OSM project, there has been an ongoing debate about the quality of the VGI. According to Basiri *et al.* (2016), the main reason for this is that many contributors to OSM have not been trained in geography or surveying and consequently their contributions, including geometry and attribute data inserts, deletions, and updates, can be inaccurate, incomplete, inconsistent, or vague. Several quality-assurance-related applications have focused on comparing OSM data with other sources of data, such as from Google Maps and Ordnance

Survey (United Kingdom) to evaluate OSM positional, temporal, and thematic accuracy and completeness of coverage.

However, it is important to note here that quality analysis of the road network (Haklay 2010), building footprints (Fan *et al.* 2014) and other features in OSM has shown satisfactory data reliability and a potential to be used in various applications. Therefore, it has been decided that this dataset is a reliable source of information for private transport network building.

The basic components of the OSM conceptual data model of the physical world consists of three types of elements: nodes (defining points in space), ways (defining linear features and area boundaries), and relations (explaining how elements work together). At the writing of this thesis there has been over 4.5 billion nodes, over 500 million ways and about 6 million relations (Open Street Map Statistics 2018).

The OSM import interface provided by the Visum software enables modeller to read the elements from the OSM database directly into Visum. Once the importing procedure is finished, the basic private transport network in Visum includes nodes, links, allowed or prohibited turns and various attributes such as number of lanes and maximum allowed speed.

The network structure imported into the PTV Visum software consists of the following elements: 18,300 nodes, 45,000 links and 126,000 turns. All these elements form the directed private transport network graph that is ready to be loaded with the estimated travel demand.

3.6.2. Public Transport Network Development

The public transport network in Visum is represented by several software specific objects that encode spatial structure, schedules and fare information. The public transport network data model is very detailed and requires at least a 10 page long description for a proper technical understanding. As this can be found in the software manual, only a concise description is given here. Generally, the network is represented by so-called stop hierarchy and line hierarchy, two concepts that are fused together (see Figure 3.6) to represent the whole public transport system.

A stop is hierarchically composed of three different elements: stop, stop area and stop point. A stop point, being the lowest object in the hierarchy, representing the geographic location where a transit vehicle physically makes its stop, i.e. a mast on the roadside most of the time. A stop area combines several stop points that are very close to each other. For example, both the northbound and southbound stop points on a particular link are grouped together as a single stop area. Stop, being the highest object in the hierarchy, encompasses all related stop areas and stop points and serves the purpose of an organizational structure rather than physical element.

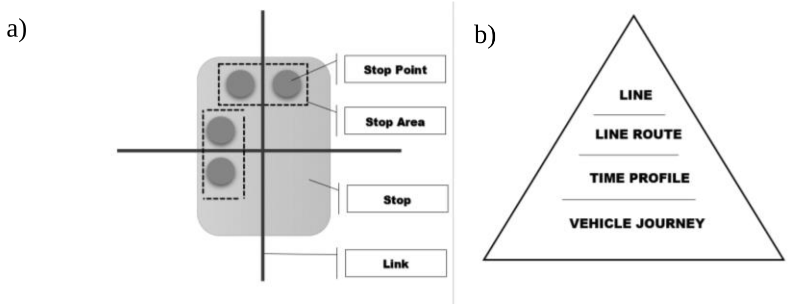

Fig. 3.6. Public Transport Network Representation (PTV Group 2014: 2798):
a) Stop Hierarchy; b) Line Hierarchy

As mentioned before, spatial routes are also hierarchically composed of the following main elements: lines, line routes, time profiles and vehicle journeys.

A line is mainly used to aggregate several line routes. The line itself neither has a spatial course in the network, nor any temporal data, i.e. run times between the stop points. The spatial structure is represented with a line route and the run times with a time profile. Specific departure times create vehicle journeys and, combined with time profiles, they describe the timetable concept.

Such a detailed representation of the public transport network requires tedious and demanding coding efforts, especially when the spatial model extent covers the whole city. Fortunately, new data sources exist now that were not present several years ago. To facilitate the network development process and to overcome a rather time-consuming task, an interface between Visum public transport network data model and General Transit Feed Specification (GTFS) data format has been utilized. Puchalsky *et al.* (2011) has employed the same methodology and recognized that such an approach: speeds up network creation and avoids manual coding errors; improves data quality over the previous approaches; and ensures an easy updating of the model information when schedules change.

According to the description given by Google (2018), GTFS defines a common format for public transportation schedules and associated geographic information. GTFS lets public transit agencies publish their transit data and lets developers write applications that consume that data in an interoperable way. A GTFS feed is composed of a series of text files (tables) collected in a ZIP (archive file format) file. Each file describes a particular aspect of transit information: stops, routes, trips, and other schedule data. The details of each file are presented in a scheme by (Wong 2013) and given in Annex C.

As can be seen from Annex C, some files (tables) and some fields are mandatory while others are optional and depend on the availability of information and an agency's willingness to publish it. The files are related to each other with shared values that are visualised as relationships in the figure above. As an example, a trip in the trip.txt file is related to a route in the route.txt file by sharing the same route identification (route_id) value. In general, the whole structure is similar to a relational database.

Since its creation, GTFS has become the most popularly used data format to describe fixed-route transit services in the world. Many agencies have decided to share their GTFS data openly with the public, while others choose to restrict access only to select partners (Antrim *et al.* 2013). In Lithuania, for example, all urban and regional bus-based public transport services are available to download as a single GTFS dataset via the website www.visimarsrutai.lt, and this initiative is supported by the Lithuanian Road Administration (2015).

In addition, it is worth recognizing other alternative GTFS applications apart from those that publish public transport information via trip planners or travel demand modelling. For example, Bok *et al.* (2016) pointed out that GTFS combined with population data is a perfect data source that allows for measuring the degree of accessibility to public transport networks and more importantly it provides an opportunity for comparative studies between different cities or metropolitan areas with different public transit systems. Karner (2018) expanded the GTFS application within the accessibility and equity field by taking into account travel times among the origins and destinations. In conclusion, the potential applicability of the data format is predicted to get even richer due to the growing set of agencies publishing data, and the ease with which the dataset can be handled.

The first Vilnius GTFS dataset was launched on 3 May 2016 by public transport operators and since then it is openly accessible online with constant updates. The dataset that was used for automatic importing into Visum was issued in February 2018 by the public transport service provider. The files comprising the dataset are described in Table 3.7.

Some of the optional GTFS files were not included into the dataset (Fare Rules, Fare Attributes, Transfers and Frequencies); however, they are not essential for modelling purposes.

Technically, the importing procedure consists of two main steps:

1. Translation of GTFS dataset into the software specific public transport system format.
2. Merging of the public transport network with the private transport network.

The first step is neither time consuming nor technically complex, as the modelling package provides an automated GTFS importer. However, the second step requires some additional manual efforts as the merging algorithm results often need to be rectified due to incorrect placement of stop points or public transport routing.

Table 3.7. Description of the General Transit Feed Specification Dataset of Vilnius Public Transport System

File	Description
Agency	The Agency file provides general information about the only public transport service provider - Municipality Enterprise “Communication Services”.
Calendar	The Calendar file defines service patterns that operate recurrently such as, for example, every weekday. There are 272 different service patterns within dataset.
Calendar Dates	The Calendar Dates table explicitly activates or disables service patterns included in Calendar file by date.
Routes	The Routes file identifies distinct routes and this concept need to be distinguished from distinct spatial routings, several of which may belong to a single route. The file describes 18 trolleybus routes, 6 night bus routes, 6 rapid bus routes and 73 conventional bus routes.
Shapes	The Shapes file describes the physical path that a vehicle takes, and all entities of this file are references within Trips table. There are 587 unique physical paths that are followed by public transport vehicles in Vilnius city.
Stops	The Stops file defines the geographic locations of each actual stop or station in the PT system. There are 1,335 different stops within the file.
Stop Times	The table provides temporal course of each public transport trip being routed via public transport stops. In total there are 24,034 service trips.
Trips	The Trips file is the central datafile within the dataset. It defines 24,034 service trips and takes up information from other files via references as can be seen in the figure above.

The importing procedure has resulted in the successful representation of the public transport network within a software data model. The GTFS dataset defines all possible public transport service trips during the course of the week, therefore it has been necessary to remove the trips taken during the weekends and retain only services related to the typical weekday. An imported public transport network model, reflecting the typical weekday 24-hour operation consists of the following items:

a) 1335 public transport stops;
b) 103 public transport lines (6 rapid bus lines, 79 conventional bus lines and 17 trolleybus lines);
c) 589 public transport line routes (34 rapid bus line routes, 422 conventional bus line routes and 133 trolleybus line routes);

d) 10,942 vehicle journeys (1,654 rapid bus vehicle journeys, 5,925 conventional bus vehicle journeys and 3,363 trolleybus vehicle journeys).

Random checks have been carried out by comparing routes and departure times with the publicly available information on websites to ensure that data has been imported without any errors and that the network model accurately represents the actual public transport network operation observed on a typical weekday.

3.7. Base and Do Something Modelling Results

It is necessary to recognize that due to the under-represented demand (see initial assumptions) and the lack of available observed data representing whole day traffic and passenger flows, there has been no way to conduct the appropriate base model route choice calibration and quality assessment within private and public transport systems. Nonetheless, morning peak hour modelled traffic flows were compared with observed on-street flows to check the deviations within the private transport route choice. The coefficient of determination (R^2) equal to 0.61 of 369 observation points show a reasonable fit. Full representation of the demand would allow more precise distribution of traffic and public transport passenger flows throught the networks; however, this is deemed as an extension to this work and remains a direction for future research.

This chapter summarizes the Base and Do Something model key outputs, i.e. statistics that define the transport system operation:

1. Vehicle Flows (VF) – The number of vehicles traversing physical infrastructure on both scenarios provided as a graphical output with numerical values.
2. Public Transport Passenger Flows (PF) – The number of passengers traversing physical infrastructure on both scenarios provided as a graphical output with numerical values.
3. Network Performance Characteristics (NPC) – Includes total vehicle kilometres, total passenger kilometres, total vehicle travel time and total passenger travel time for both scenarios.

Outputs given and compared for both scenarios in this section represent the whole day, however more finely tuned outputs are also accessible and can be examined upon interest. The following two figures (see Figure 3.7 and Figure 3.8) present the modelled passenger car flows (vehicles per day) for the area of interest, i.e. in close proximity to the planned Siaurine Street.

Fig. 3.7. Base Model Daily Vehicle Flows (Created by Author)

The figure above presents base model daily vehicle flows. It is evident that Geležinio Vilko, Ukmergės and T. Narbuto Streets is the backbone structure serving the bulk of daily demand.

Fig. 3.8. Do Something Daily Vehicle Flows (Created by Author)

It can be concluded that the new arterial street will serve between 11,000 to 12,000 vehicle trips of the modelled demand in both directions per day. It is worth noting here that the modelled demand is lower than the real one (as noted in initial assumptions section) and, thus, the actual flows after implementation are expected to be higher to some extent. The further two figures (see Figure 3.9 and Figure 3.10) define the modelled daily public transport passenger flows for the area of interest.

Fig. 3.9. Base Model Daily Public Transport Passenger Flows (Created by Author)

Fig. 3.10. Do Something Model Daily Passenger Flows (Created by Author)

Apparently, the impact on public transport flows is very negligible as the two drawings above look almost identical with only small variations in passenger flows.

The overal network performance measures allows for a holistic perspective on the infrastructure development incentives. Table 3.8 given below provides the quantitative comparison of Base and Proposed models in terms of overall network performance characteristics.

Table 3.8. Base and Do Something Model's Network Daily Performance Characteristics

Performance Measure	Base	Do Something	Absolute Difference	Relative Difference
Total number of person trips by public transport	274,900	268,400	–6,500	–2.4%
Mean person journey time by public transport, minutes	32.0	32.0	0	0.0%
Total journey time by public transport, hours	162,500	162,200	–300	–0.2%
Total number of passenger kilometres	1791,400	1790,000	–1,400	–0.1%
Total number of person trips by car	888,800	898,700	9,900	1.1%
Mean person journey time by car, minutes	25.4	25.2	–0.2	–0.8%
Total journey time by car, hours	385,000	382,000	–3,000	–0.8%
Total number of people kilometers by car	9262,000	9253,000	–9,000	–0.1%

Due to improved traffic conditions (quicker trips) and the appearance of new convenient routes, the total number of car users is expected to increase by 9,900. These users are attracted from competing public transport and walk modes. Even though there will be some reduction in total people travel time (3,000 hours) and total number of people kilometres (9,000 km), the average trip by car duration will be reduced only marginaly from 25.4 minutes to 25.2 minutes. This reduction would have been much more significant if additional users had not been attracted.

Because of improved traffic conditions, public transport is expected to experience some negative impact. It is predicted that there will be a loss of 6,500 users and consequently total passenger kilometres and total journey time will decrease.

3.8. Travel Demand Model Development Framework

Based on the literature review, and practical experience gained through empirical data collection and model development, the summarizing travel demand model development framework is proposed and discussed within this section. In essence, the framework consists of three main chronological stages:

1. First stage: Data collection.
2. Second stage: Model development.
3. Third stage: Model calibration and quality assessment.

A detailed description of the processes that are supposed to be undertaken with their descriptions are given in Table 3.9 that follows.

Table 3.9. Travel Demand Model Development Methodological Framework

Stage	Process	Detailed Description
1.1	Definition of Spatial and Temporal Scope	1.1.1. Definition of Spatial Model Extents and Transport Analysis Zones. 1.1.2. Retrieval of Datasets Describing Population Distribution and Economic Activities within Transport Analysis Zones. 1.1.3.Definition of Modelled Time Period and Resolution.
1.2	Administration of Travel Behaviour Survey	1.2.1. Questionnaire Design and Sampling Strategy. 1.2.2. Data Collection. 1.2.3. Identification of activity sequences and their probabilities.
1.3	Identification of External Demand	1.3.1. Identification of External Zones. 1.3.2. Data Collection. 1.3.3. Identification of tours originating at external zones.
1.4	Traffic and Passenger Data Collection	1.4.1. Identification of Relevant Network Movements. 1.4.2. Collection of Traffic and Public Transport Passenger Flows. 1.4.3. Data Analysis and Presentation.
2.1	Network Model Development	2.1.1. Retrieval of Relevant Open Street Map and General Transit Feed Specification Databases. 2.1.2. Transfer of the Retrieved Data to PTV Visum Data Structures.
2.2	Generation of Matrices for External Zones	2.2.1. Conversion of External Demand Data into the Matrix Format. 2.2.2. Expansion of Resulting Matrices to Reflect the Full External Flows
2.3	Tour-based Travel Demand Model Definition	2.3.1. Encoding of Population Segments and their Distributions.

End of Table 3.9

Stage	Process	Detailed Description
		2.3.2. Encoding of Workplace (Attractiveness) Distributions. 2.3.3. Specification of Destination Choice and Mode Choice Sub Models.
3.1	Calibration and Quality assessment of Tour Generation and Mode Choice Procedures	3.1.1. Calibration and Quality Assessment of Destination Choice and Mode Choice Models.
3.2	Calibration and Quality Assessment of Route Choice Procedure	3.2.1. Private and Public Transport Assignment to Networks. 3.2.2. Private and Public Transport Route Choice Calibration and Quality Assessment.

First two stages are linear and a process comprising them idealy should be undertaken in consecutive order, although exceptions apply. The third stage is iterative and the processes (3.1 and 3.2) should be repeated in turn until satisfactory quality assessment results are achieved.

This framework was implemented almost to its fullest extent in this thesis. Unfortunately, some processes (1.3, 2.2 and 3.1) were not followed due to non-availability of data and the extensive financial requirements necessary to collect it.

3.9. Conclusions of the Third Chapter

1. This 3rd chapter has focused on the development of the tour-based travel demand model, which relies heavily on the various datasets that can be either purposefully surveyed or harnessed from open source datasets. An accurate 24-hour transport supply representation has been achieved by employing open source datasets such as Open Street Map and General Transit Feed Specification Data. A typical trip-based travel demand model deals with partial temporal (2 separate peak periods with 1-hour scope) transport supply representation.
2. The demo tour-based model can be deemed as fit for purpose. Quality assessment of destination choice compared modelled and observed average trip lengths for all the distinct trip purposes and the model was calibrated to ensure no higher discrepancies than 5%. In addition, quality assessment of the mode choice compared modelled and observed modal shares within a 5 km travel distance bands. The model was calibrated to ensure no higher discrepancies than 8%. The comparison of modelled and

observed morning peak hour traffic flows shows reasonable traffic replication ($R^2 = 0.61$, n = 369).

3. The tour-based model was employed in the assessment of urban transport network development scenario i.e. a foreseen implementation of Siaurine Street. A comparison of Base and Do Something scenarios reveals a reduction in the average one trip travel time from from 25.4 min/trip to 25.2 min/trip. Due to the additional car users (9,900) attracted from public transport (6500) and to some extent walk mode (3500), the total number of vehicle kilometres is going to decrease only by 9,000 km/day (–0.1%), whereas the total number of vehicle hours by 3,000 hours/day (–0.8%). This is expected to negatively impact the performance of public transport system as the decreased ridership (6,500 persons per day) is expected to reduce revenues.
4. Only a part of the overall travel demand is represented within this tour-based demand model. In reality, however, there are additional components that generate it. In the current model version, no demand is generated at the external zones due to the lack of available data. In addition, internal zones do not account for the demand generated by movement of goods and movement of population members younger than 16-year-old. To estimate demand for the youngest population segment, the travel behaviour need to be identified, whereas to account for the freight traffic generated at the internal zones, potentially a separate tour-based freight travel demand model can be developed and added to the current model version.

General Conclusions

1. The trip-based model treats person's trips as independent decisions bringing in significant errors in the modelling results. Tour-based models can resolve this issue as they deal with the sequences of trips (a.k.a. tours) and consider their interlocking dependencies. Tour-based travel demand model is superior to trip-based travel demand model, because of the up to four times larger set of covered trip purposes, a capability to take into account daily trip sequences and a consistent mode choice within the tour.
2. An innovative activity sequence-focused survey of travel behavior and associated data analysis methodology proposed within the scope of this thesis allows identification of daily activity sequences and their probabilities for the homogeneous population segments. With this method, insights about travel distance and travel time estimates is gained through secondary datasets coming from Google travel time database and associated route choice algorithms. As a result, the errors of travel time and distance estimates are eliminated.
3. Open source and datasets play a significant role within the transport model development exercise. An application of these datasets within the transport modelling process ensures precise daily travel supply representation, and reduces the financial cost. Open Street Map dataset was trans-

ferred into the model environment to represent the private transport network graph geometry and associated property data with the utmost spatial precision. Similarly, General Transit Feed Specification dataset was utilized to realistically represent public transport routes and schedules.

4. The demo model developed for urban area can be deemed as fit for purpose. Quality assessment of destination choice compared modelled and observed average trip lengths for all the distinct trip purposes and the model was calibrated to ensure no higher discrepancies than 5% across all trip purposes. In addition, quality assessment of the mode choice compared modelled and observed modal shares within a 5 km travel distance bands. The model was calibrated to ensure no higher discrepancies than 8% across all distance bands. Route choice calibration and quality assessment have not been conducted due to the under-represented demand. However, the comparison of modelled and observed morning peak hour traffic flows shows reasonable traffic replication ($R^2 = 0.61$, $n = 369$). The full demand representation and the achievement of a rigorous route choice is deemed to be a direction for further research.
5. The tour-based model developed as part of the thesis was applied to assess the impact of Siaurine Street construction scenario to network performance. It is forecasted that the total number of vehicle kilometres will decrease by 9,000 km/day (–0.1%), whereas the total number of vehicle hours by 3,000 hours/day (–0.8%).
6. Based on the outcomes of this research, a comprehensive transport modelling framework is created**.** The framework defines the structure and sequencing of the three main stages: data collection, model development and model quality assessment. The first two stages are sequential, whereas the third one is of an iterative nature. This modelling framework can serve as the best practice guide for the applied and scientific studies to ensure statistical reliability, transparency and methodological consistency.

References

de Abreu e Silva J. 2018. The Effects of Land-use Patterns on Home-based Tour Complexity and Total Distances Traveled: A Path Analysis, *Sustainability (Switzerland)* 10(3). DOI: 10.3390/su10030830.

Allström A., Kristoffersson I., Susilo Y. 2017. Smartphone-based Travel Diary Collection: Experiences from a Field Trial in Stockholm, *44th European Transport Conference in Barcelona, Spain*. Elsevier B.V. DOI: 10.1016/j.trpro.2017.07.006.

Antrim A., Barbeau S. J. 2013. The Many Uses of GTFS Data – Opening the Door to Transit and Multimodal Applications, *ITS America. Nashville, Tennessee.*

Arsanjani J. J., Mooney P., Zipf A. 2015. Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets *Open Street Map in GIScience*. DOI: 10.1007/978-3-319-14280-7.

Axhausen K. W. 2007. Definition of Movement and Activity for Transport Modelling, *Handbooks in Transport*. Elsevier, Oxford.

Banister D. 2008. The Sustainable Mobility Paradigm, *Transport Policy* 15(2): 73–80. DOI: 10.1016/j.tranpol.2007.10.005.

Basiri A., Jackson M., Amirian P., Pourabdollah A., Sester M., Winstanley A., Moore T., Zhang L. 2016. Quality Assessment of OpenStreetMap Data Using Trajectory Mining, *Geo-Spatial Information Science* 19(1): 56–68. DOI: 10.1080/10095020.2016.1151213.

Bellei G., Gentile G., Papola N. 2005. A Within-day Dynamic Traffic Assignment Model for Urban Road Networks, *Transportation Research Part B: Methodological* 39: 1–29. DOI: 10.1016/j.trb.2004.02.002.

Ben-Akiva M., Bierlaire M. 1999. Discrete Choice Methods And Their Applications To Short Term Travel Decisions, *Handbook of Transportation Science. International Series in Operations Research & Management Science*. Springer, Boston.

Ben-Akiva M., Lerman S. R. 1994. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT press, Cambridge.

Blumenstock J., Maldeniya D., Lokanathan S. 2017. Understanding the Impact of Urban Infrastructure: New Insights from Population-Scale Data, *Proceedings of the 9th International Conference on Information and Communication Technologies and Development - ICTD '17*: 1–12. DOI: 10.1145/3136560.3136575.

Bok J., Kwon Y. 2016. Comparable Measures of Accessibility to Public Transport using the General Transit Feed Specification, *Sustainability (Switzerland)* 8(3). DOI: 10.3390/su8030224.

Carriere D. 2018. Geocoder: API Documentation [cited 4 March 2017] . Available online: http://geocoder.readthedocs.io.

Cascetta E., Nuzzolo A., Russo F., Vitetta A. 1996. A Modified Logit Route Choice Model Overcoming Path Overlapping Problems. Specification and Some Calibration Results for Interurban Networks. *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*. Elsevier Science, Lyon, France.

UK Department for Transport. 2014a. *TAG UNIT M3.1 Highway Assignment Modelling* [Cited 18 May 2018]. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/427124/webtag-tag-unit-m3-1-highway-assignment-modelling.pdf

UK Department for Transport. 2014b. *TAG UNIT M3.2 Public Transport Assignment* [Cited 20 May 2018]. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/427126/webtag-tag-unit-m3-2-public-transport-assignment-modelling.pdf

Dijkstra E.W. 1959. A Note on Two Problems in Connexion with Graphs, *Numerische Mathematik* 1(1): 269–271. DOI: 10.1007/BF01386390.

Duranton B. G., Turner M. A. 2011. Fundamental Law of Road Congestion: Evidence from U.S. Cities, American Economic Review 101(6): 2616–2652. DOI: 10.1257/aer.101.6.2616.

European Parliament, Council of the European Union. 2013. Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information, *Official Journal of the European Union*: 1–8 [cited 10 July 2018]. Available online: http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF.

Fan H., Zipf A., Fu Q., Neis P. 2014. Quality Assessment for Building Footprints Data on Open Street Map, *International Journal of Geographical Information Science* 28(4): 700–719. DOI: 10.1080/13658816.2013.867495.

Fellendorf M., Haupt T., Heidl U., Scherr W. 1997. Activity-based Demand Forecasting in Daily Practice. *Activity-based Approaches to Travel Analysis*. Elsevier Science. Oxford. United Kingdom.

Fellendorf M., Haupt T., Heidl U., Scherr W. 2000. *VISEM - An Activity Chain-based Traffic Demand Model*, Karlsruhe.

Friedrich M. 1998. A Multi-Modal Transport Model for Integrated Planning. *8th World Conference on Transport Research*. Pergamon. Antwerp.

Friedrich M. 2008. Multi-day Dynamic Transit Assignment, *Schedule-based Modeling of Transportation Networks* 46*: 1-11*. DOI: 10.1007/978-0-387-84812-9_3.

Friedrich M., Hofsäß I., Wekeck S. 2001. Timetable-based Transit Assignment Using Branch & Bound Techniques, *Transportation Research Records* (1752): 100–107. DOI: 10.3141/1752-14.

Friedrich M., Wekeck S. 2002. A Schedule-based Transit Assignment Model Addressing the Passengers ' Choice among Competing Connections. *The Schedule-based Approach in Dynamic Transit Modelling: Theory and Applications* 28:159–173. DOI: 10.1007/978-1-4757-6467-3_9.

Goodchild M. F. 2007. Editorial : Citizens as Voluntary Sensors : Spatial Data Infrastructure in the World of Web 2 . 0. *International Journal of Spatial Data Infrastructures Research* 2: 24–32. DOI: 10.1016/j.jenvrad.2011.12.005.

Google LCC. 2018. General Transit Feed Specification [cited 15 June 2018]. Available online: https://developers.google.com/transit/gtfs/reference?csw=1.

Google LLC. 2018. Google Maps Distance Matrix API: Developer's Guide [cited 25 February 2017]. Available online: https://developers.google.com/maps/documentation/distance-matrix/intro#Introduction.

Gorham R. 2009. *Demystifying Induced Travel Demand*, Bonn, Germany.

Grigonis V. 2018. *Analysis and Modelling of Siaurine Street Taking into Account Traffic Quality and Safety*, Vilnius.

Habibian M., Kermanshah M. 2011. Exploring the role of transportation demand management policies' interactions, *Scientia Iranica* 18(5): 1037–1044. DOI: 10.1016/j.scient.2011.09.005.

Haklay M. 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets, *Environment and Planning B: Planning and Design* 37(4): 682–703. DOI: 10.1068/b35097.

Hensher D. A., Reyes A. J. 2000. Trip Chaining as a Barrier to the Propensity to Use Public Transport, *Transportation* 27(4):341. DOI: 10.1023/A:1005246916731.

Hernsher D. A., Rose J. M., Greene W.H. 2005. *Applied Choice Analysis. A Primer*. New York: Cambridge University Press.

Horni A., Nagel K., Axhausen K.W. 2016. *The Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press. DOI: http://dx.doi.org/10.5334/baw.

Institute for Transport Studies. 2017. German Mobility Panel. Available online: https://www.ifv.kit.edu/english/26_MOP.php.

Karner A. 2018. Assessing Public Transit Service Equity using Route-level Accessibility Measures and Public Data, *Journal of Transport Geography* 67: 24–32. DOI: 10.1016/j.jtrangeo.2018.01.005.

Kepaptsoglou K., Meerschaert V., Neergaard K., Papadimitriou S., Rye T., Schremser R., Vleugels I. 2012. Quality Management in Mobility Management: A Scheme for Supporting Sustainable Transportation in Cities, *International Journal of Sustainable Transportation* 6(4): 238–256. DOI: 10.1080/15568318.2011.587137.

Krizek K. J. 2003. Neighborhood services, trip purpose, and tour-based travel, *Transportation* 30(4): 387–410. DOI: 10.1023/A:1024768007730.

Land A. H., Doig A. G. 1960. An Automatic Method for Solving Discrete Programming Problems, *Econometrica* 28(3): 497–520. DOI: 10.2307/1910129.

Lee K. S., Eom J. K., Moon D. S. 2014. Applications of TRANSIMS in transportation: A literature review, *Procedia Computer Science* 32: 769–773. DOI: 10.1016/j.procs.2014.05.489.

Lee M. S., Chung J. H., McNally M. G. 2002. An Empirical Investigation of the Underlying Behavioural Processes of Trip Chaining, *University of California Transportation Center, Working Papers.*

Levinson D., Kumar A. 1995. Activity, Travel, and the Allocation of Time, *Journal of the American Planning Association* 61(4): 458–470. DOI: 10.1080/01944369508975657.

Lithuanian Road Administration. 2015. GTFS Datasets [cited 15 June 2018]. Available online: http://www.visimarsrutai.lt/gtfs/.

Litman T. 2003. The Online TDM Encyclopedia: Mobility management information gateway, *Transport Policy* 10(3): 245–249. DOI: 10.1016/S0967-070X(03)00025-8.

Loidl M., Wallentin G., Cyganski R., Graser A., Scholz J., Haslauer E. 2016. GIS and Transport Modeling – Strengthening the Spatial Perspective, *ISPRS International Journal of Geo-Information* 5(6): 84. DOI: 10.3390/ijgi5060084.

Ma J., Mitchell G., Heppenstall A. 2014. Daily Travel Behaviour in Beijing , China : An Analysis of Workers ' Trip Chains , and the Role of Socio-demographics and Urban Form, *Habitat International* 43: 263-273. DOI: 10.1016/j.habitatint.2014.04.008.

McFadden D. 1974. *Conditional Logit Analysis of Qualitative Choice Behaviour*. New York: Academic Press.

McGuckin N., Zmud J., Nakamoto Y. 2005. Trip-Chaining Trends in the United States: Understanding Travel Behaviour for Policy Making, *Transportation Research Record* 1917: 199–204. DOI: 10.3141/1917-22.

McKinney W. 2008. Pandas Documentation [cited 18 March 2018].. Available online: http://pandas.pydata.org/.

McKinney W. 2013. *Python for Data Analysis* [cited 15 March 2018].. Available online: http://shop.oreilly.com/product/0636920023784.do.

McNally M. G., Rindt C. R. 2007. The Activity-based Approach, *Handbook of Transport Modelling* 1: 55–73.

Ministry of Communication. 2015. *Sustainable Urban Mobility Guidelines* [cited 15 March 2017]. Available online:https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/a1c919e0c9cc11e4bc22872d979254dd?jfwid=-l5uh8xdz3.

Ministry of Environment. 2015. *Planning Guidelines for Urban Transport Systems* [cited 14 March 2017]. Available online: http://www.am.lt/VI/files/File/TPUAD/TP%20nauji/Normu_projektas%2001%2016.pdf

Municipality Enterprise Communication Services. 2018. Vilnius Public Transport Timetables and Route Information in GTFS format [cited 12 May 2018] . Available online: http://www.stops.lt/vilnius/vilnius/gtfs.zip.

Nagel K., Beckman R., Barrett C. 1999. TRANSIMS for Urban Planning, *6th International Conference on Computers in Urban Planning and Urban Management*.

Oliphant T. 2006. NumPy: Array Processing for Numbers, Strings, Records, and Objects [cited 15 February 2018]. Available online: www.numpy.org.

Omer M., Kim H., Sasaki K., Nishii K. 2010. A tour-based travel demand model using person trip data and its application to advanced policies, *KSCE Journal of Civil Engineering* 14(2): 221–230. DOI: 10.1007/s12205-010-0221-6.

Open Data Commons. 2017. Open Database License [cited 2 May 2018]. Available online: https://opendatacommons.org/licenses/odbl/1.0/.

Open Street Map Contributors. 2018. Open Street Map [cited 3 May 2018]. Available online: https://www.openstreetmap.org.

Open Street Map Statistics. 2018. Open Street Map Daily Statistics [cited 3 May 2018]. Available online: https://www.openstreetmap.org/stats/data_stats.html.

OpenStreetMap Wiki Contributors. 2016. OpenStreetMap Wiki [cited 5 May 2018]. Available online: https://www.openstreetmap.org.

Ortuzar D. de J., Willumsen L.G. 2011. *Modelling Transport.* 4th Edition. New Delhi: A John Wiley and Sons.

PTV Group. 2014. PTV Visum 17 – Manual.

Puchalsky C., Joshi D., Scherr W. 2011. Development of a Regional Forecasting Model Based on Google Transit Feed, *91st Annual Meeting of the Transportation Research Board*, Washington.

QGIS. 2018. QGIS User Guide [cited 17 March 2018]. Available online: http://docs.qgis.org/2.8/en/docs/index.html.

Reichman S. 1976. Travel Adjustments and Life Styles: a Behavioural Approach, *Behavioural Travel-Demand Models*. Lexington.

Rossi T. F., Shiftan Y. 1997. Tour-based Travel Demand Modeling in the U.S., *IFAC Proceedings Volumes* 30(8): 381–386. DOI: 10.1016/S1474-6670(17)43853-5.

Santos G., Maoh H., Potoglou D., von Brunn T. 2013. Factors Influencing Modal Split of Commuting Journeys in Medium-Size European Cities, *Journal of Transport Geography* 30: 127–137. DOI: 10.1016/j.jtrangeo.2013.04.005.

Schmiedel R. 1984. Determination of Behaviour-oriented Groups of People for Transport Planning. Karlsruhe.

Shafique M. A., Hato E. 2015. A Review of Travel Data Collection Methods, *17th International Conference on Civil Engineering and Applied Mechanics at Paris, France.*

State Enterprise Centre of Registers. 2018. Commercial Entities Spatial Database [cited 7 February 2018]. Available online: https://www.registrucentras.lt/.

Strathman J. G., Dueker K. J., Davis J. S. 1994. Effects of Household Structure and Selected Travel Characteristics on Trip Chaining, *Transportation* 21(1): 23–45. DOI: 10.1007/BF01119633.

The Department of Statistics. 2015. Official Statistics Portal [cited 2 March 2018]. Available online: https://osp.stat.gov.lt

The Department of Statistics. 2011. Census Data 2011 [cited 7 February 2018]. Available online: https://osp.stat.gov.lt.

Train K. 1993. *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand.* Cambridge: The MIT Press.

Transportation Research Board. 2018. Travel Forecasting Resource [cited 2 February 2018]. Available online: http://tfresource.org.

Victoria Transport Policy Institute. 2003. Online TDM Encyclopedia [cited 28 January 2018]. Available online: http://www.vtpi.org/tdm/.

Vilniaus Planas. 2018. Sustainable Urban Mobility in Vilnius [cited 20 July 2018] Available online: http://judumas.vilnius.lt/vdjp-informacija/.

Vovsha P., Bradley M., Bowman J. L. 2005. Activity-based Travel Forecasting Models in the United States: Progress since 1995 and Prospects for the Future, *Progress in Activity-based Analysis.* Oxford, Elsevier.

Wallace B., Barnes J., Rutherford G. S. 2000. Evaluating the effects of traveller and trip characteristics on trip chaining, with implications for Transportation Demand Management Strategies, *Journal of the Transportation Research Board* 1718:97–106. DOI: 10.3141/1718-13.

Wang F., Xu Y. 2011. Estimating O-D travel time matrix by Google Maps API: implementation, advantages, and implications, *Annals of GIS* 17(4): 199–209. DOI: 10.1080/19475683.2011.625977.

Wilson A. G. 1967. A Statistical Theory of Spatial Distribution Models, *Transportation Research* 1(3): 253–269. DOI: https://doi.org/10.1016/0041-1647(67)90035-4.

Wong J. C. 2013. Leveraging the General Transit Feed Specification for Efficient Transit Analysis, *Transportation Research Record: Journal of the Transportation Research Board* 2338: 11–19. DOI: 10.3141/2338-02.

Ye X., Pendyala R. M., Gottardi G. 2007. An Exploration of the Relationship between Mode Choice and Complexity of Trip Chaining Patterns, *Transportation Research Part B: Methodological* 41(1): 96–113. DOI: 10.1016/j.trb.2006.03.004.

Zheng H., Son Y. J., Chiu Y. C., Head L., Feng Y., Xo H., Kim S., Hickman M. 2013. A Primer for Agent-based Simulation and Modeling in Transportation Applications [cited 20 March 2018]. Available online: https://www.fhwa.dot.gov/publications/research/ear/13054/13054.pdf

Zhong M., Shan R., Du D., Lu C. 2015. A Comparative Analysis of Traditional Four-Step and Activity-based Travel Demand Modeling : A Case Study of Tampa, Florida, *Transportation Planning and Technology* 38(5): 517–533. DOI: 10.1080/03081060.2015.1039232.

List of Scientific Publications by the Author on the Topic of the Dissertation

Articles in the Reviewed Scientific Journals

Grigonis, V; Burinskienė, M; Paliulis, G; Ušpalytė-Vitkūnienė, R; Dumbliauskas, V; Barauskas, A. Modelling a Passenger Car System based on the Principles of Sustainable Mobility in Vilnius City. 2014. *Transport* 29(3): 334–341 (Clarivate Analytics Web of Science).

Dumbliauskas, V; Grigonis, V; Vitkienė, J. Estimating the Effects of Public Transport Priority Measures at Signal Controlled Intersections. 2017a. *The Balic Journal of Road and Bridge Engineering* 12(3):187–192 (Clarivate Analytics Web of Science).

Dumbliauskas, V; Grigonis, V; Barauskas, A. Application of Google-based Data for Travel Time Analysis – Kaunas City Case. 2017b. *PROMET – Traffic&Transportation*. 29(6): 613–621 (Clarivate Analytics Web of Science).

Dumbliauskas, V; Grigonis, V; Barauskas, A. The Impact of Cost-Benefit Analysis on Decision Making Concerning the Development of the Urban Transport System: Case of Kaunas City.2018. *Transport* 33(4) 2018: 334–341 (Clarivate Analytics Web of Science).

Articles in Other Editions

Dumbliauskas, V; Barauskas, A. Analysis of Kaunas City Population and Workplace Density in Terms of Mobility. 2015. *Science – Future of Lithuania* 7(5):528–532.

Summary in Lithuanian

Įvadas

Problemos formulavimas

Lietuvos susisiekimo sistemų planavimo metodika sparčiai vystėsi nuo nepriklausomybės atkūrimo 1990 metais. Pagrindinis norminis dokumentas, reglamentuojantis projektavimą ir planavimą iki 2015 metų buvo Aplinkos ministerijos statybos techninis reglamentas „Miestų, miestelių ir kaimų susisiekimo sistemos", kuris buvo performuotas į du atskirus norminius dokumentus. Šiuo metu statybos techninis reglamentas „Gatvės ir vietinės reikšmės keliai" reglamentuoja infrastruktūros projektavimą, o „Urbanizuotų teritorijų susisiekimo sistemų planavimo normos" apibrėžia miesto susisiekimo sistemų planavimo procesą. Susisiekimo ministerija, su Europos Komisijos iniciatyva ir parama, taip pat ėmėsi skatinti darnesnį miesto gyventojų judumą. Šiuo tikslu buvo patvirtintos Darnaus judumo mieste planų rengimo gairės detalizuojančios Darnaus judumo planų turinį, pačią planavimo procedūrą ir suinteresuotų šalių vaidmenį. Minėti metodiniai ir transporto sistemos prioritetinių krypčių plėtotės pokyčiai ateityje neišvengiamai turės teigiamų pokyčių planavimo praktikai ir pačiai susisiekimo sistemai.

Vis dėlto naujas metodinių dokumentų rinkinys, nesudaro deramo pagrindo objektyviam ir analitiniais metodais grįstam sprendimų priėmimui. „Urbanizuotų teritorijų susisiekimo sistemų planavimo normose" yra teigiama, jog rengiant miestų, kurie turi daugiau kaip 50 tūkstančių gyventojų, bendruosius planus turi būti atliktas susisiekimo poreikių modeliavimas. Tačiau toks teiginys neužtikrina deramo rezultato, nes nėra aiškios ir visuotinai naudojamos modelių taikymo metodikos.

Darbo aktualumas

Dėl paminėtų priežasčių didžioji dauguma itin brangių miesto infrastruktūros plėtros sprendimų dažniausiai yra grindžiami inžinerine intuicija ar politiniais poreikiais. Visgi, siekiant užtikrinti visuomenės poreikius ir racionalų finansinių išteklių panaudojimą, miesto susisiekimo sistemos planavimas turi būti atliekamas taikant analitinius metodus, tokius kaip susisiekimo poreikių modeliavimas ir naudos bei kaštų analizė. Pavyzdžiu gali būti laikoma Jungtinė Karalystė įteisinusi išsamų gairių rinkinį, kuris turi gilias tradicijas ir yra plačiai taikomas praktikoje.

Šia disertacija siekiama pasiūlyti metodologiją, kuri būtų pakankamai paremta gerąja praktika ir pažangiomis technologijomis ir kuri taip pat būtų lengvai taikoma.

Tyrimo objektas

Tiriamojo darbo objektas yra miesto susisiekimo sistemos komponentai: susisiekimo paklausa ir susisiekimo pasiūla.

Darbo tikslas

Šiuo darbu siekiama pasiūlyti pažangią susisiekimo poreikių modelio rengimo metodiką, kuri įvertintų visos dienos susisiekimo pasiūlą ir paklausą.

Darbo uždaviniai

Darbo tikslui pasiekti buvo suformuluoti šie uždaviniai:

1. Sugretinti ir palyginti šiuo metu praktikoje priimtas taikyti ir teoriniame lygmenyje vystomas susisiekimo poreikių modelių rengimo metodikas.
2. Suformuoti ir patikrinti efektyvias ir inovatyvias duomenų rinkimo procedūras, kurios įvertina visos dienos susisiekimo pasiūlą ir paklausą.
3. Sukurti susisiekimo poreikių modelį įvertinantį visos dienos susisiekimo paklausą ir pasiūlą, bei panaudoti jį susisiekimo tinklo plėtotės scenarijų vertinimui.
4. Suformuoti apibendrinančią susisiekimo poreikių modelio rengimo metodiką, kurios pagalba būtų rengiami statistiškai patikimi modeliai ir kuri būtų lengvai taikoma susisiekimo tinklo planavimui.

Tyrimų metodika

Siekiant darbo tikslo buvo pritaikytos įvairios plačiai pripažintos metodikos ir įrankiai:

1. Metodiška mokslinės literatūros apžvalga leido nustatyti stipriuosius ir silpnuosius susisiekimo modelių rengimo metodikų aspektus.

2. Empirinė keliavimo įpročių apklausa ir antrinių kelionės atstumo duomenų rinkimas taikant „Google Distance Matrix API“ buvo panaudoti tiriant respondentų judumo parametrus.
3. „Python“ programavimo kalba ir jos standartinės bibliotekos („Pandas“, „NumPy“, „SciPy“, „Seaborn“) buvo panaudota išgavimui, apdorojimui ir statistinei analizei duomenų, naudojamų transporto pasiūlos ir paklausos aprašymui susisiekimo poreikių modelyje.
4. Atviro kodo geografinė informacinė sistema QGIS buvo panaudota atliekant geoerdvinės analizės užduotis ir ruošiant duomenis susisiekimo pasiūlos ir paklausos aprašymui susisiekimo poreikių modelyje.
5. „PTV Visum“ modeliavimo aplinka buvo panaudota susisiekimo paklausos ir pasiūlos duomenų valdymui, analizei ir modeliavimui.

Darbo mokslinis naujumas

Darbo naujumas apibūdinamas šiais aspektais:

1. Suformuota inovatyvi empirinių keliavimo įpročių duomenų rinkimo, apdorojimo ir analizės metodika leidžianti nustatyti kelionių sekas ir jų tikimybes homogeniškose populiacijos grupėse. Ši inovatyvi keliavimo įpročių tyrimo metodika gali būti naudinga atliekant keliavimo įpročių analizę mokslinėse ir taikomosiose studijose.
2. Kuriant susisiekimo pasiūlos modelį taikomi nauji ir atviri duomenų rinkiniai (OSM ir GTFS). Tai užtikrina tikslų susisiekimo tinklo aprašymą modelyje ir mažesnes šio modelio kūrimo sąnaudas.
3. Sukurtas ir praktiškai pritaikytas inovatyvus kelionių grandinėmis pagrįstas susisiekimo poreikių modelis panaudojant šiuolaikiškas duomenų išgavimo ir apdorojimo technologijas.
4. Suformuluota išsami apibendrinanti susisiekimo poreikių modelių kūrimo metodika, kurią visuotinai taikant, būtų užtikrintas modeliavimo rezultatų patikimumas ir metodinis vientisumas.

Darbo rezultatų praktinė reikšmė

Inovatyvūs duomenų apdorojimo metodai pritaikyti ruošiant susisiekimo paklausos ir transporto pasiūlos duomenis kuriamam kelionių grandinėmis pagrįstam modeliui. Modelio pagalba galima vertinti daugybę kitų susisiekimo tinklo vystymo (naujos viešojo transporto rūšys, „Statyk ir važiuok“ sistema ir pan.) ir susisiekimo poreikių valdymo (naujos urbanizuojamos teritorijos, veiklų centrų perskirstymas ir pan.) scenarijų. Kelionių grandinėmis pagrįstas modelis pritaikytas vertinant Šiaurinės gatvės įtaką susisiekimo poreikių pasiskirstymui ir transporto tinklo funkcionavimui.

Apibendrinanti susisiekimo poreikių modelių kūrimo metodika yra universali ir pritaikoma ne tik Lietuvos miestams, bet ir miestams užsienyje. Visgi papildomos duomenų rinkimo, apdorojimo, analizės ir modelio rengimo pastangos yra neatsiejama proceso dalis.

Ginamieji teiginiai

1. Kelionių grandinėmis pagrįsti susisiekimo poreikių modeliai yra pranašesni nei atskiromis kelionėmis pagrįsti modeliai, kadangi kelionių grandinėmis pagrįsti modeliai įvertina platesnę kelionių aibę, visos dienos kelionių grandines ir nuoseklų susisiekimo būdo pasirinkimą grandinėse.
2. Nuoseklus ir reguliariais intervalais pasikartojantis veiklų sekomis pagrįstos metodikos taikymas gyventojų keliavimo įpročių analizei panaudojant pažangias technologijas užtikrintų galimybę retrospektyvinei analizei, detaliai atskleistų keliavimo įpročius homogeniškoms populiacijos grupėms, ir leistų aprašyti visos dienos susisiekimo paklausą susisiekimo poreikių modeliuose.
3. Atvirų duomenų rinkinių panaudojimas susisiekimo paklausos ir pasiūlos aprašymui modeliuose užtikrintų labai tikslų viešojo ir privataus susisiekimo tinklų aprašymą ir finansinės projekto kainos sumažinimą.
4. Apibendrinanti susisiekimo poreikių modelių kūrimo metodika yra esminis elementas statistiškai patikimam ir nuosekliam susisiekimo poreikių modelių kūrimui. Teikiama apibendrinanti susisiekimo poreikių modelių rengimo metodika gali būti traktuojama kaip gerosios praktikos vadovas užtikrinantis patikimas modelių charakteristikas.

Darbo rezultatų aprobavimas

Disertacijos tema atliktų tyrimų rezultatai buvo publikuoti penkiuose straipsniuose. Keturi straipsniai buvo publikuoti recenzuojamuose mokslo žurnaluose, kurie yra reitinguojami *Clarivate Analytics Web of Science* duomenų bazėje. Vienas straipsnis buvo publikuotas mokslo žurnale kuris reitinguojamas kitose duomenų bazėse:

Mokslinių tyrimų rezultatai buvo paskelbti dvejose tarptautinėse mokslinėse konferencijose:

1. Barauskas, A; Dumbliauskas, V. A Study on the Travel Time Generalized Cost Function in Kaunas City. 2016. 28th European Conference of Operational Research (Poznan University of Technology).
2. Dumbliauskas, V; Grigonis, V. Estimating the Effects of Public Transport Priority Measures at Signal Controlled Intersections. 2017. 10th International Conference of Environmental Engineering (VGTU).

Disertacijos struktūra

Disertaciją sudaro trys pagrindiniai skyriai, kurie glaustai gali būti apibūdinti taip: susisiekimo poreikių modelių kūrimo metodikų apžvalga, susisiekimo poreikių empirinis tyrimas (metodika, duomenų rinkimas ir statistinė analizė) bei kelionių grandinėmis pagrįsto modelio kūrimas.

Disertaciją sudaro 126 puslapiai, 90 literatūros šaltinių, 32 paveikslai, 21 lentelės ir 43 numeruotos formulės.

1. Susisiekimo poreikių modelių kūrimo metodikų apžvalga

Šiuo metu dėmesio centre yra trijų tipų susisiekimo poreikių modeliai, kurie pagal vystymosi chronologiją išsidėsto taip:

1. Tradiciniai kelionėmis pagrįsti susisiekimo poreikių modeliai (angl. *trip-based models*).
2. Kelionių grandinėmis pagrįsti susisiekimo poreikių modeliai (angl. *tour-based models*).
3. Individualių asmenų sprendimų seką simuliuojantys modeliai (angl. *agent-based models*).

Kelionėmis pagrįsti modeliai, kurių pirmieji prototipai buvo sukurti XX amžiaus viduryje tapo industrijos *de facto* standartu ir iki šiol yra plačiai taikomi daugelyje išsivysčiusių šalių. Kelionėmis pagrįsto modelio principiniai komponentai yra keturi:

1. Kelionių generavimas (angl. *trip generation*).
2. Kelionių paskirstymas (angl. *trip distribution*).
3. Susisiekimo būdo parinkimas (angl. *mode choice*).
4. Susisiekimo poreikių priskyrimas susisiekimo tinklui (angl. *assignment*).

Vis dėlto nuo šio modelio atsiradimo pradžios buvo pastebėti keli reikšmingi jo trūkumai:

a) modelyje kelionių pradžios ir pabaigos agreguojamos į transportines zonas, o tai sukuria reikšmingas modelio paklaidas;
b) modelyje kelionės traktuojamos kaip nepriklausomos, nesusijusios tarpusavyje, o tai turi neigiamą įtaką kuriamų susisiekimo poreikių matricų reprezentatyvumui;
c) modelis neturi galimybės įvertinti įvairių judumo valdymo iniciatyvų (pvz. įvažiavimo į miesto centrinę dalį apmokestinimo) įtakos susisiekimo sistemos naudotojų keliavimo įpročiams;
d) modelyje tik išimtinais atvejais yra galimybė vertinti dinaminius susisiekimo poreikių ir susisiekimo pasiūlos svyravimus.

Visuotinai pripažinti modelio trūkumai ir augantys modeliavimo ekspertų poreikiai atlikti inovatyvias analizes privertė mokslininkus peržvelgti esamą situaciją ir ieškoti alternatyvių pažangesnių metodų (Rossi *et al.* 1997: 381–386).

Kaip alternatyva, praktikams buvo pasiūlyti kelionių grandinėmis pagrįsti modeliai (Fellendorf *et al.* 2000), kurie aprašo keliones ne kaip nepriklausomus vienetus, tačiau, kaip tarpusavyje susijusias kelionių grandines, gyventojų atliekamas paros metu. Kelionių grandinėmis pagrįsti modeliai susideda iš šių pagrindinių žingsnių:

a) veiklų grandinių generavimas (angl. *generation of activity sequences*);
b) kelionių grandinių generavimas ir susisiekimo būdų parinkimas kelionių grandinių kelionėms (angl. *generation of tours and estimation of mode choice*);
c) susisiekimo poreikių priskyrimas susisiekimo tinklui (angl. *assignment*).

Šią procedūrą lyginant su klasikiniu modeliu, galima identifikuoti esminius skirtumus ties pirmais dviem etapais. Empirinio tyrimo duomenų pagalba pirmajame etape generuojamos veiklų grandinės, o antrame – kiekvienai veiklai priskiriamos transportinės zonos, t. y. geografinės koordinatės. Nors ši procedūra neišsprendžia visų klasikinio modelio trūkumų, tačiau joje kelionės yra tarpusavyje susijusios, todėl kelionių paskirstymas ir susisiekimo būdo parinkimo etapai labiau priartėja prie realybės.

Kelionių grandinėmis pagrįstas modelis yra tinkama alternatyva nacionalinėje praktikoje taikomiems klasikiniams modeliams, iki kol individualių asmenų sprendimų seką simuliuojantys (angl. *agent-based models*) modeliai pasieks brandos stadiją. Ši metodika testuojama „PTV Visum" aplinkoje, kurios pagalba kuriamas modelis Vilniaus miestui.

2. Susisiekimo poreikių empirinis tyrimas: metodika, duomenų rinkimas ir statistinė analizė

Šiame skyriuje teikiama empirinio miesto gyventojų judumo įpročių tyrimo metodika ir tyrimo rezultatų analizė. Minėto tyrimo tikslas yra nustatyti ir kiekybiškai išreikšti šias miesto gyventojų judumo ypatybes:

a) vidutinį kelionių skaičių;
b) kelionių pasiskirstymą tarp susisiekimo būdų;
c) kelionių pasiskirstymą tarp keliavimo tikslų;
d) vidutinį kelionių ilgį;
e) veiklų pradžios laikus;
f) veiklų grandines ir jų tikimybes.

Visi šie parametrai yra svarbūs kelionių grandinėmis pagrįsto modelio kūrimui. Sukurta apklausa ir jos atlikimo metodika pritaikyta tiriant Vilniaus miesto gyventojų populiacijos judumo charakteristikas. Apklausa buvo siekiama nustatyti šios populiacijos narių mobilumą apibūdinančių parametrų reprezentatyviąsias reikšmes. Būtų pravartu, jei tokios apklausos būtų rengiamos reguliariai, pavyzdžiui, kas penkerius metus. Tokiu būdu būtų sudaroma galimybė atlikti retrospektyvinę analizę ir atnaujinti susisiekimo poreikių modelį. Pavyzdžiui Vokietijoje tokias apklausas kas penkeri metai organizuoja Karlsrūjės technologijos instituto padalinys finansuojant Susisiekimo ir skaitmeninės infrastruktūros ministerijai. Vilniaus miesto gyventojų apklausos atlikimo planavimas buvo įvykdytas pagal šiuos žingsnius:

1. Pilnos potencialių respondentų aibės nustatymas.
2. Apklausos metodo parinkimas.
3. Respondentų atrankos strategijos parinkimas.
4. Imties dydžio nustatymas.

Dėl teisinių ir apklausos atlikimo terminų apribojimų buvo nuspręsta galimų respondentų aibę apriboti neįtraukiant jaunesnių kaip šešiolikos metų asmenų. Apklausa buvo atlikta pasitelkiant du metodiškai skirtingus žingsnius:

1. Pirmuoju žingsniu (2017 metų birželio ir liepos mėnesiais), atrinkti respondentai buvo paprašyti užpildyti anketas internetu.
2. Antruoju žingsniu respondentai, kurie neužpildė anketų internetu, buvo aplankyti namuose (2017 metų rugsėjo mėnesį) ir apklausti gyvai.

Respondentų atranka buvo atliekama taikant sluoksninės atrankos metodą:

1. Pirmame etape visa imtis buvo paskirstyta tarp transportinių zonų proporcingai pagal juose gyvenančių žmonių skaičių.
2. Antrajame etape respondentai buvo atsitiktinai atrenkami iš transportinėse zonose gyvenančių žmonių.

Apklausos imtį sudarė 1773 respondentai arba 0,33 % visų Vilniaus miesto gyventojų. Empirinio tyrimo rezultatai rodo, jog respondentai atlieka vidutiniškai po 2,3 keliones per dieną, o kiekviena kelionė yra 7,5 km ilgio. 21 % kelionių yra atliekama viešuoju transportu, 61 % automobiliu ir 15 % pėsčiomis. Dažniausi kelionių tikslai yra darbovietė (24 %), apsipirkimo vieta (11 %) ir namai (38 %).

Didžioji dalis apklaustųjų (45,3 %) atlieka paprastąsias kelionių grandines, tai yra atlieka vieną veiklą ir keliauja atgal namo. Sudėtingas kelionių grandines atlieka 30,8 % apklaustų respondentų, o 23,9 % neatlieka kelionių apskritai.

Labiausiai tikėtina paprastoji veiklų grandinė yra „Namai–Darbas–Namai" atliekama su 0,296 tikimybe, tuo tarpu labiausiai tikėtina sudėtingoji veiklų grandinė yra „Namai–Darbas–Apsipirkimas–Namai" atliekama su 0,049 tikimybe. Veiklų grandinės ir jų tikimybės teikiamos S2.1 lentelėje.

S2.1 lentelė. Nustatytos veiklų grandinės ir jų tikimybės

Veiklų grandinė	Stebėjimų skaičius	Santykinis dažnis	Atlikimo tikimybė
Visos veiklos atliekamos namuose	424	21,4 %	0,296
MWM	525	26,5 %	0,239
MSM	122	6,2 %	0,069
MWSM	86	4,3 %	0,049
MHM	48	2,4 %	0,027
MCM	46	2,3 %	0,026
MEM	42	2,1 %	0,024
MLM	38	1,9 %	0,021
MXM	37	1,9 %	0,021
MOM	29	1,5 %	0,016
MBM	26	1,3 %	0,015
MWBM	26	1,3 %	0,015
MAM	25	1,3 %	0,014
MDM	25	1,3 %	0,014
MDWPM	22	1,1 %	0,012
MWCM	18	0,9 %	0,010
MDWM	16	0,8 %	0,009
MPM	12	0,6 %	0,007
MWAM	11	0,6 %	0,006
MWLM	10	0,5 %	0,006
MWXM	10	0,5 %	0,006
MKM	8	0,4 %	0,005
MBSM	7	0,4 %	0,004

S2.1 lentelės pabaiga

Veiklų grandinė	Stebėjimų skaičius	Santykinis dažnis	Atlikimo tikimybė
MCSM	7	0,4 %	0,004
MHSM	7	0,4 %	0,004
MSSM	7	0,4 %	0,004
MWASM	6	0,3 %	0,003
MWBSM	6	0,3 %	0,003
MWPM	6	0,3 %	0,003
MWSCM	6	0,3 %	0,003
Visos kitos grandinės	321	16,2 %	0,181
Suma:	1979	100,0 %	1,116

Veiklų sekų tikimybės yra apskaičiuojamos nustatytą sekų skaičių dalijant iš imties dydžio. Dalis apklaustųjų teigė nepaliekantys namų visą dieną (įvairius darbus atliekantys namuose), todėl lentelėje pateikiama papildoma eilutė, nurodanti tikimybę, jog respondentas lieka namuose visą dieną.

Ši inovatyvi keliavimo įpročių tyrimo metodika gali būti naudinga atliekant keliavimo įpročių analizę ir kitose mokslinėse bei taikomosiose studijose.

3. Kelionių grandinėmis pagrįsto modelio kūrimas ir šio modelio taikymo rezultatų analizė

Šioje disertacijos dalyje aprašomas kelionių grandinėmis pagrįsto susisiekimo poreikių modelio kūrimas, naudojant „PTV Visum“ aplinką. Šiame etape yra itin reikšmingi pirmajame skyriuje aprašyti metodai ir antrajame skyriuje identifikuoti keliavimo įpročius aprašantys parametrai.

Erdvės atžvilgiu parengtas modelis apima Vilniaus miesto ir priemiestinės teritorijas su bendru 282 transportinių zonų kiekiu. Laiko atžvilgiu modelis įvertina vienos tipinės dienos susisiekimo poreikius su vienos valandos dinamine rezoliucija. Modelyje aprašytas visos teritorijos kelių/gatvių tinklas ir Vilniaus miesto viešojo transporto sistema. Demonstraciniais tikslais, modelis taikomas vertinant planuojamos Šiaurinės gatvės įtaką susisiekimo poreikiams ir miesto susisiekimo sistemos funkcionavimui.

Susisiekimo poreikių modelio kūrimo etapai:

1. Populiacijos suskirstymas į homogeniškas grupes.
2. Veiklų grandinių formavimo procedūros aprašymas.
3. Veiklų paskirstymo dienos eigoje aprašymas.
4. Kelionių grandinių formavimo procedūros aprašymas.
5. Susisiekimo būdo parinkimo procedūros aprašymas.

Modeliuojama gyventojų populiacija (Vilniaus miestas ir priemiestinės teritorijos) skaidoma į aštuonis segmentus, kurių viduje keliavimo įpročiai laikomi homogeniškais, t. y. taikomos vienodos kelionių grandinės ir jų tikimybės. Skaidymui naudojami du kriterijai: amžius su trimis kategorijomis (16–19 metų, 20–49 metai, virš 50 metų) ir užimtumas su keturiomis kategorijomis (studentai, dirbantys, nedirbantys, pensininkai). Tolimesniems modelio kūrimo etapams yra būtina turėti gyventojų segmentų dydžius visose modelio transportinėse zonose. Šiam tikslui naudojami Lietuvos statistikos departamente kaupiami erdviniai gyventojų pasiskirstymo duomenys, kurie apdorojami ir parengiami GIS įrankiais.

Veiklų grandinių generavimo procedūros pagalba įvertinamas bendras kiekvieno populiacijos segmento judumo poreikis kiekvienoje transportinėje zonoje. Šiame etape segmentų dydžiai transportinėse zonose dauginami iš empiriškai nustatytų veiklų grandinių tikimybių. Skaičiavimo rezultatas (kiekvieno populiacijos segmento veiklų grandinių dydis) yra naudojamas kituose modelio žingsniuose.

Siekiant, kad susisiekimo poreikiai būtų aprašyti vienai tipinei dienai su vienos valandos dinamine rezoliucija, reikia įvertinti veiklų pasiskirstymą dienos bėgyje. Šiame žingsnyje skaičiavimai atliekami kiekvienam veiklos tipui priskiriant apklausos metu empiriškai nustatytą tikimybinį pasiskirstymo dienos bėgyje skirstinį.

Kelionių grandinių generavimo procedūra priskiria geografines koordinates (transportines zonas) visoms veikloms ir tokiu būdu joms suteikia erdvinę dimensiją, o taip pat kartu sukuria poreikį judėti tarp transportinių zonų. Šiame etape taikomas diskrečiųjų pasirinkimų „Logit“ tikimybinis modelis, kuris naudoja zonų patrauklumo ir zonų erdvinės atskirties duomenis. Zonų patrauklumas išreiškiamas veiklai aktualių darbo vietų tankiu, o erdvinė atskirtis – kelionės atstumo ir laiko funkcijomis.

Susisiekimo būdo parinkimo procedūra priskiria vieną iš trijų modeliuojamų sistemų (keliavimas viešuoju transportu, asmeniniu automobiliu ar pėsčiomis) kiekvienai grandinės kelionei.

Šiam žingsniui taip pat taikomas diskrečiųjų pasirinkimų „Logit“ tikimybinis modelis. Skirtingai nuo atskiromis kelionėmis pagrįstų modelių, kelionių grandinėmis pagrįstuose modeliuose šis žingsnis padeda įvertinti nuoseklų susisiekimo būdo parinkimą visoje grandinėje. Pavyzdžiui, jeigu pirmoji grandinės kelionė atliekama automobiliu, tuomet visos grandinės kelionės yra atliekamos automobiliu.

Susisiekimo poreikių modelyje susisiekimo tinklo aprašymas atliekamas pritaikant atvirai prieinamus „Open Street Map“ ir „General Transit Feed Specification“ duomenis. Šių duomenų panaudojimas reikšmingai pagerina modeliuojamų susisiekimo poreikių ir susisiekimo tinklo reprezentatyvumą bei sumažina bendrą projekto kainą. „Open Street Map“ duomenų rinkmena buvo konvertuota į susisiekimo poreikių modelio aplinką taip sukuriant tikslią gatvių ir kelių grafo geometriją bei aprašomųjų duomenų rinkmeną. Konvertuota „General Transit Feed Specification“ duomenų rinkmena leido tiksliai reprezentuoti viešojo transporto sistemos maršrutus ir tvarkaraščius.

Kryptinis gatvių ir kelių tinklo grafas susideda iš 18300 mazgų ir 45000 atkarpų. Viešojo transporto sistema reprezentuojama 1335 stotelėmis, 103 viešojo transporto maršrutais, 589 viešojo transporto maršrutų linijomis ir 10942 viešojo transporto priemonių kelionėmis atliekamomis per tipinę darbo dieną.

Sukurtas demonstracinis kelionių grandinėmis pagrįstas modelis yra vidutiniškai tikslus. Vertinant veiklų atlikimo vietos parinkimo skaičiavimo procedūros kokybę buvo palygintas vidutinis modeliuojamų ir apklausos būdu nustatytų kelionių ilgis. Kalibruojant modelį buvo užtikrintas ne didesnis kaip 5 % skirtumas visose veiklų/kelionių grupėse. Vertinant susisiekimo būdo parinkimo skaičiavimų kokybę buvo palyginti susisiekimo būdo skirstiniai 5 kilometrų kelionių ilgio diapazonuose. Kalibruojant modelį buvo užtikrintas ne didesnis kaip 8 % skirtumas visuose kelionės ilgio diapazonuose. Dėl nepilnai reprezentuotų susisiekimo poreikių nebuvo galimybės atlikti kelionių priskyrimo tinklui skaičiavimo procedūros kalibravimo. Visgi preliminarus rytinio piko privataus transporto modeliuojamų ir gatvėje stebimų srautų palyginimas rodo vidutinišką modelio tikslumą (R^2 = 0,61, n = 369). Pilnas susisiekimo poreikių reprezentavimas galėtų būti atliekamas toliau tęsiant tyrimą.

S3.1 lentelė. Esamos ir planuojamos būklės susisiekimo tinklo funkcionavimo charakteristikos

Charakteristika	Esama būklė	Planuojama būklė	Absoliutus skirtumas	Santykinis skirtumas
Viešuoju transportu atliekamų kelionių skaičius, vnt.	274900	268400	–6500	–2,4 %
Vidutinis vienos viešuoju transportu atliekamos kelionės laikas, min	32,0	32,0	0	0,0 %
Bendras kelionių viešuoju transportu laikas, h	162500	162200	–300	–0,2 %
Bendras viešojo transporto keleivių nukeliautas atstumas, km	1791400	1790000	–1400	–0,1 %
Privačiu automobiliu atliekamų kelionių skaičius, vnt.	888800	898700	9900	1,1 %
Vidutinis vienos kelionės automobiliu laikas, min	25,4	25,2	–0,2	–0,8 %
Bendras kelionių automobiliu laikas, h	385000	382000	–3000	–0,8 %
Bendras automobiliais nukeliautas atstumas, km	9262000	9253000	–9000	–0,1 %

Sukurtas kelionių grandinėmis pagrįstas modelis panaudotas vertinant susisiekimo tinkle vystymo scenarijaus, t. y. planuojamos Šiaurinės gatvės statybos įtaką susisiekimo poreikių pasiskirstymui ir tinklo funkcionavimui. Eksperimentinio taikymo rezultatai teikiami S3.1 lentelėje. Pagal vidutinę vienos kelionės lengvuoju automobiliu trukmę, matyti, kad bendras gatvių tinklo funkcionavimas pagerėtų nuo 25,4 min/vienai kelionei iki 25,2 min/vienai kelionei. Nors pagerėjusios privataus susisiekimo sąlygos pritrauks papildomus vartotojus (apie 9900), kurie naudojasi kitais susisiekimo būdais, vis dėlto bendra privačiu automobiliu atliekamų kelionių rida sumažės 9000 km per dieną (–0,1 %), o bendras privačių automobilių kelionės laikas 3000 valandų (–0,8 %). Dėl šių priežasčių, viešojo transporto sistema praras dalį vartotojų (6500 per dieną) ir tai neigiamai atsilieps jos funkcionavimui.

Remiantis atliktu moksliniu tyrimu, darbe suformuluota apibendrinanti susisiekimo poreikių modelių kūrimo metodika teikiama S3.2 lentelėje.

S3.2 lentelė. Apibendrintoji susisiekimo poreikių modelių rengimo metodika

Nr.	Procesas	Proceso aprašymas
1.1	Modelio apimties nustatymas	1.1.1 Modelio geografinės apimties nustatymas ir transportinių zonų kūrimas.
		1.1.2 Duomenų apie populiaciją ir darbo vietas transportinėse zonose rinkimas.
1.2	Gyventojų keliavimo įpročių identifikavimas	1.2.1 Apklausos atlikimo planavimas.
		1.2.2 Duomenų rinkimas.
		1.2.3 Apklausos duomenų analizė ir veiklų sekų identifikavimas.
1.3	Susisiekimo poreikių išorinėse zonose nustatymas	1.3.1 Išorinių zonų identifikavimas ir apklausos planavimas.
		1.3.2 Duomenų rinkimas.
		1.3.3 Kelionių grandinių identifikavimas.
1.4	Transporto ir keleivių srautų tyrimai	1.4.1 Aktualių manevrų tinkle identifikavimas.
		1.4.2 Transporto ir keleivių srautų duomenų rinkimas.
		1.4.3. Duomenų apdorojimas ir analizė.
2.1	Susisiekimo tinklo modelio kūrimas	2.1.1 OSM ir GTFS duomenų rinkmenų įsigijimas.
		2.1.2 Duomenų rinkmenų konvertavimas į „PTV Visum“ aplinkos duomenų struktūras.
2.2	Išorinių susisiekimo poreikių aprašymas	2.2.1 Išorinių susisiekimo poreikių perkėlimas į matricinį formatą.
		2.2.2 Ištirtų išorinių poreikių išplėtimas.
2.3	Vidinių susisiekimo poreikių aprašymas	2.3.1 Populiacijos segmentų apibrėžimas ir pasiskirstymo duomenų perkėlimas į transportines zonas.
		2.3.2 Darbo vietų skirstinio perkėlimas į transportines zonas.
		2.3.3 Kelionių grandinių generavimo ir susisiekimo būdo parinkimo modelių aprašymas.
3.1	Kelionių grandinių generavimo procedūros kalibravimas ir kokybės vertinimas	3.1.1 Kelionių grandinių generavimo ir susisiekimo būdo parinkimo modelių parametrų kalibravimas ir kokybės vertinimas.
3.2	Priskyrimo tinklui procedūros kalibravimas ir kokybės vertinimas	3.2.1 Transporto ir keleivių srautų priskyrimas tinklui.
		3.2.2 Transporto ir keleivių srautų priskyrimo tinklui procedūros kalibravimas ir kokybės vertinimas.

Metodikoje apibrėžiama užduočių struktūra ir atlikimo seka, kuri sudaryta iš trijų pagrindinių etapų: duomenų rinkimas (1), modelio kūrimas (2) ir modelio kalibravimas bei kokybės vertinimas (3). Pirmieji du etapai rengiami nuosekliai, tuo tarpu trečiasis etapas atliekamas iteraciniu būdu. Ši metodika gali būti gerosios praktikos vadovas mokslinėse ir taikomosiose studijose siekiant užtikrinti modeliavimo rezultatų patikimumą, skaidrumą ir metodinį vientisumą.

Bendrosios išvados

1. Kelionėmis pagrįstas susisiekimo poreikių modelis aprašo asmenines keliones kaip tarpusavyje nepriklausomus vienetus, tačiau toks susisiekimo poreikių aprašymas nulemia reikšmingas modeliavimo rezultatų paklaidas. Kelionių grandinėmis pagrįstas modelis išsprendžia šią problemą aprašydamas asmenines keliones kaip tarpusavyje susijusių sprendimų seką. Kelionių grandinėmis pagrįstas susisiekimo poreikių modelis yra pranašesnis, kadangi šis modelis įvertina iki keturių kartų platesnę kelionių tipų aibę ir visos dienos metu atliekamų kelionių grandines.
2. Disertacijoje pasiūlyta inovatyvi veiklų grandinėmis pagrįsta gyventojų keliavimo įpročių apklausos kūrimo, atlikimo ir jos duomenų analizės metodika leidžia nustatyti visos dienos veiklų grandines ir jų tikimybes homogeniškoms populiacijos grupėms. Nukeliauto atstumo ir kelionės trukmės vertinimas metodikoje atliekamas panaudojant antrinius Google kelionės trukmės duomenis ir maršruto planavimo algoritmus. Šis metodas užtikrina kelionės atstumo ir trukmės vertinimo paklaidų eliminavimą.
3. Atviri duomenys yra esminė susisiekimo poreikių modelių kūrimo dalis. Šių duomenų panaudojimas užtikrina tikslų visos dienos susisiekimo pasiūlos reprezentatyvumą. „Open Street Map“ duomenų rinkmena buvo konvertuota į susisiekimo poreikių modelio aplinką taip sukuriant tikslią gatvių ir kelių grafo geometriją bei aprašomųjų duomenų rinkmeną. Pertvarkyta „General Transit Feed Specification“ duomenų rinkmena leido tikroviškai aprašyti viešojo transporto sistemos maršrutus ir tvarkaraščius.
4. Sukurtas visos dienos demonstracinis kelionių grandinėmis pagrįstas modelis yra vidutiniškai tikslus. Vertinant veiklų atlikimo vietos parinkimo skaičiavimo procedūros kokybę buvo palygintas vidutinis modeliuojamų ir apklausos būdu nustatytų kelionių ilgis. Kalibruojant modelį buvo užtikrintas ne didesnis kaip 5 % skirtumas visose veiklų/kelionių grupėse. Vertinant susisiekimo būdo parinkimo skaičiavimų kokybę buvo palyginti susisiekimo būdo skirstiniai 5 kilometrų kelionių ilgio diapazonuose. Kalibruojant modelį buvo užtikrintas ne didesnis kaip 8 % skirtumas visuose kelionės ilgio diapazonuose. Preliminarus rytinio piko privataus transporto modeliuojamų ir gatvėje stebimų srautų palyginimas rodo vidutinišką modelio tikslumą ($R^2 = 0{,}61$, $n = 369$).
5. Sukurtas kelionių grandinėmis pagrįstas modelis panaudotas vertinant Šiaurinės gatvės statybos scenarijaus įtaką susisiekimo tinklo funkcionavimui. Prognozuojama, kad gyvendinus projektą bendra privačiu automobiliu atliekamų kelionių rida sumažės 9000 km per dieną (–0,1 %), o bendras privačių automobilių kelionės laikas 3000 valandų (–0,8 %).

6. Remiantis atliktu moksliniu tyrimu, darbe suformuluota apibendrinanti susisiekimo poreikių modelių kūrimo metodika. Metodikoje apibrėžiama užduočių struktūra ir atlikimo seka, kuri sudaryta iš trijų pagrindinių etapų: duomenų rinkimas, modelio kūrimas ir modelio kokybės vertinimas. Pirmieji du etapai rengiami nuosekliai, tuo tarpu trečiasis etapas atliekamas iteraciniu būdu. Ši metodika gali būti gerosios praktikos vadovas mokslinėse ir taikomosiose studijose siekiant užtikrinti modeliavimo rezultatų statistinį patikimumą, skaidrumą ir metodinį vientisumą.

Annexes[2]

Annex A. Distribution of Trips by Trip Start Time and Activity

Annex B. The Most Likely Activity Sequences and their Probabilities

Annex C. General Transit Feed Specification Data Model

Annex D. Author's Declaration of Academic Integrity

Annex E. The Co-Authors' Agreements to Present the Material of Publications as a Part of the Doctoral Dissertation

Annex F. Copies of Scientific Publications by the Author on the Topic of the Dissertation

[2] The annexes are provided in the enclosed compact disc.

Vytautas DUMBLIAUSKAS

DEVELOPMENT AND APPLICATION OF TOUR-BASED TRAVEL DEMAND MODEL FOR PLANNING OF URBAN TRANSPORT NETWORKS

Doctoral Dissertation

Technological Sciences,
Transport Engineering (T 003)

KELIONIŲ GRANDINĖMIS PAGRĮSTO SUSISIEKIMO POREIKIŲ MODELIO KŪRIMAS IR TAIKYMAS MIESTŲ SUSISIEKIMO TINKLO PLANAVIMUI

Daktaro disertacija

Technologijos mokslai,
transporto inžinerija (T 003)

2019 03 08. 10,75 sp. l. Tiražas 20 egz.
Vilniaus Gedimino technikos universiteto
leidykla „Technika“,
Saulėtekio al. 11, 10223 Vilnius,
http://leidykla.vgtu.lt
Spausdino BĮ UAB „Baltijos kopija“,
Kareivių g. 13B, 09109 Vilnius