A Time-Space Constrained Approach for Modeling Travel and Activity Patterns

Dissertation

zur Erlangung des akademischen Grades
Dr. rer. nat. im Fach Geographie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
der Humboldt-Universität zu Berlin

von
Diplom Geograph Andreas Justen

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Gutachter/in:
1. Prof. Dr. Barbara Lenz
2. Prof. Francisco Javier Martínez Concha, PhD
3. PD Dr. Christian Schiller

eingereicht: 04.05.2011
Tag der Verteidigung: 14.07.2011

Acknowledgement

At first, I would like to thank my supervisor Prof. Francisco Martínez (University of Chile, Santiago, Division of Transport Engineering) for his encouragement, his support and patience. I am grateful for his generous expert guidance through this ‘scientific journey’, that is the PhD. I owe thanks to Prof. Dr. Barbara Lenz for being my supervisor and the opportunity to develop the study at the DLR Institute of Transport Research, Berlin. She always offered me advice and supported this project and my professional career. Special thanks to PD Dr. Christian Schiller (University of Dresden, Department Theory of Transportation Planning) for valuable discussions and accepting the appraisal of the thesis. Special thanks are due also to Prof. Cristián Cortés (University of Chile, Santiago) who gave me sound advice during innumerable working sessions in Santiago and in Berlin.

I also wish to extend my gratitude to all my colleagues at the DLR Institute of Transport Research. In particular I appreciated the stimulating discussions and the professional company of my colleague Rita Cyganski in the past years. Many thanks are devoted to Katja Köhler for her advice on statistics and to Stefan Trommer for our conversations and his sympathy also at tougher moments.

My appreciation is deemed to the colleagues from the Division of Transport Engineering at the University of Chile who were generous with their time, insight, and hospitality during my working stays in Chile. Felipe Sanhueza provided me with information on Santiago models of transport and land use. Alan Thomas and Esteban Godoy from Sectra (Secretaría de Planificación de Transporte) in Santiago offered support and willingly provided information.

The PhD was initiated during the course of the research initiative ‘Risk Habitat Megacity’, funded by the Helmholtz-Association. Thus I am grateful to the project managers – in person to Dirk Heinrichs and Kerstin Krellenberg – for the opportunity to be part of the project and participate in the activities organized to provide and advance exchange between the Chilean and German PhD students.

Mein ganz besonderer Dank gilt meinen Eltern Wolfgang und Christiane, die mich seit jeher voll und ganz in meinen Ideen und Vorhaben unterstützen. Vielen Dank für Eure motivierende Zusprache und das grenzenlose Vertrauen. Danke an meine Schwester Anja und Heiko für viele befreiende Treffen, an denen es gerade nicht um die Doktorarbeit ging. Ja, und ganz vorne stehen bei dieser Danksagung müsste eigentlich Ana. Gracias por tu apoyo ilimitado y tu capacidad de hacer que este trabajo haya sido una experiencia sumamente positiva para mí.

ACKNOWLEDGEMENT 2

LIST OF TABLES 5

LIST OF FIGURES 6

LIST OF FORMULAS 7

ZUSAMMENFASSUNG 8

ABSTRACT 9

1 INTRODUCTION 10
1.1 Objectives and approach 11
1.2 Chapter structure 13

2 A REVIEW OF TRANSPORT DEMAND MODELS 14
2.1 Trip-based approaches 14
2.2 Trip-chaining approaches 18
2.3 Activity-based approaches 22
2.4 Constraints and the role of time geography 26
2.5 Summary 29

3 METHODOLOGY FOR THE MODELING OF ACTIVITY AND TRAVEL PATTERN 32
3.1 The hierarchical structure of choices 32
3.1.1 The concept of macro and micro processes 35
3.1.2 Hierarchies of decision-making in daily travel 37
3.1.3 The hierarchical tour – terminology 39
3.2 The hierarchical tour – methodology 41
3.3 Research questions 50

4 DATA AND MODELS: EVIDENCES FROM SANTIAGO 52
4.1 Santiago transport and land-use models 52
4.2 EOD: Santiago Travel and Household Survey 56
4.3 Activity and travel behavior in Santiago 56
4.3.1 Summary on main data preparation steps 57
4.3.2 Activity prioritization 59
4.3.3 Tours and activity patterns 66
4.4 Summary 71

5 EMPIRICAL ANALYSIS OF TIME-SPACE CONSTRAINTS 72
5.1 Dependency of mode choices 73
5.2 Total daily travel time 76
5.3 Daily activity times 77
5.3.1 Grouping of activity starting times and duration 77
5.3.2 Influence factors on time regimes 81

5.4 Detour factors ____ 84
5.4.1 Conceptual background ____ 84
5.4.2 Empirical detour factors ____ 87
5.5 Local attraction ____ 94
5.5.1 Attractiveness ____ 95
5.5.2 Tour relations ____ 96
5.5.3 Accessibility ____ 98
5.6 Summary ____ 100

6 MODEL APPLICATION AND RESULTS ____ 102
6.1 Estimation of pattern demand ____ 103
6.2 The hierarchical choice process ____ 105
6.2.1 Defining the search space ____ 105
6.2.2 The daily travel time restriction ____ 107
6.2.3 Matching level: search spaces and EOD ____ 109
6.2.4 Ranking of secondary activities locations ____ 110
6.2.5 The final choice set ____ 112
6.3 The spatial path flow ____ 117
6.4 Calibration of destination choice ____ 119
6.4.1 Scenario: Adjusted travel times ____ 123
6.4.2 Visualization of spatial path flow ____ 127
6.5 Time regimes: EOD and model ____ 128
6.6 Reduction of choice options: quantification ____ 129

7 DISCUSSION AND OUTLOOK ____ 132
7.1 Transferability of the approach ____ 133
7.2 The magical number seven ____ 135
7.3 Suggestions for further development ____ 135
7.4 Closing remarks ____ 138

REFERENCES ____ 140

ANNEX ____ 145
Annex-1: Aggregation of travel purposes ____ 145
Annex-2: Mode combination probabilities ____ 150
Annex-3: Final choice set ____ 151
Annex-4: Time regimes: duration of primary activity ____ 152
Annex-5: Daily travel times by mode combinations: model and EOD ____ 153

DIGITAL ANNEX ____ 154

List of tables

Table 3.1: Time hierarchies for activities 34
Table 4.1: Household user groups by income and car ownership 55
Table 4.2: Primary activities durations 63
Table 4.3: Typology of activity patterns 69
Table 5.1: Daily travel time by mode combination 76
Table 5.2: Correlation between time regimes, user groups and city sectors 82
Table 5.3: Detour factors by mode combination 88
Table 5.4: Detour factors by primary activity duration 89
Table 5.5: Detour factors by user group 89
Table 5.6: Detour factors by distance between home and work 89
Table 5.7: Influence of distance and time on detour factors - parameter estimates 92
Table 5.8: Detour factors by time periods and distance categories 94
Table 5.9: Share of home and work related tours by mode combinations 97
Table 5.10: OD travel times by mode and time slots 98
Table 6.1: Activity pattern types probabilities by user group 104
Table 6.2: Trip demand for pattern type ‚HWYH' 105
Table 6.3: Level of matching: EOD vs. modeled search spaces 109
Table 6.4: Example for the definition of the number of choice alternatives 113
Table 6.5: Exemplary results of the model 118
Table 6.6: EOD results – trip distributions by mode to secondary activity locations 119
Table 6.7: Model-1 results – trip distributions by mode, LWPs = 1 120
Table 6.8: Model-1 results – trip distributions by mode, LWPs after iterations 121
Table 6.9: Comparison of travel times – EOD and ESTRAUS 123
Table 6.10: Travel time adjustment factors 124
Table 6.11: Model-2 results – trip distributions by mode, LWPs = 1 125
Table 6.12: Model-2 results – trip distributions by mode, LWPs after iterations 125

Annex-Table 1: Temporal characteristics of travel purposes in the EOD 145
Annex-Table 2: Modal split by travel purposes in the EOD 146
Annex-Table 3: Distance matrix between EOD travel purposes 147
Annex-Table 4: Mode combination probabilities by user groups 150
Annex-Table 5: Final choice set selection by mode combinations and tour relations 151
Annex-Table 6: Overview of SPSS-Syntax used for EOD data mining 154
Annex-Table 7: Overview of SPSS-Syntax to run the model 154
Annex-Table 8: Overview of the model's input and output data files 155

List of figures

Figure 1.1: Chapter structure....13
Figure 2.1: Scheme of the four-step algorithm....17
Figure 2.2: Concepts of tours (trip-chains) and activity pattern....19
Figure 3.1: Macro and micro scale: sequence of choice processes....38
Figure 3.2: Methodology – calculation procedure....49
Figure 4.1: Maps of spatial levels in Santiago....52
Figure 4.2: Interaction between land-use and transport models....54
Figure 4.3: Input and output of Santiago’s land-use and transport models....54
Figure 4.4: Example of geocoded EOD survey information....56
Figure 4.5: Example of an EOD daily activity pattern....57
Figure 4.6: Data mining: reduction steps....58
Figure 4.7: Comparison of EOD data set - before and after data mining....59
Figure 4.8: Activity prioritization: two-level hierarchy....62
Figure 4.9: Activity prioritization: three-level hierarchy....64
Figure 4.10: Activity prioritization: activity duration....65
Figure 4.11: Frequency distribution of activity pattern in Santiago....67
Figure 4.12: Shares of grouped activity pattern types....70
Figure 5.1: Share of mode combinations, pattern type ‘HXYH’....73
Figure 5.2: Tour-based modal split values / (N=758)....74
Figure 5.3: Mode choice combinations by household user groups....75
Figure 5.4: Starting time and duration for the primary activity....80
Figure 5.5: Duration of the secondary activity....80
Figure 5.6: Starting hour of primary activity by city sector and user groups....81
Figure 5.7: Distance measures....84
Figure 5.8: Illustration of the detour concept....85
Figure 5.9: Examples: minimum convex polygon and standard deviation ellipse....85
Figure 5.10: Search space given the detour factor....87
Figure 5.11: Percentage increase of detour factors....91
Figure 5.12: Santiago Metropolitan Region – density of commerce and service land-use....95
Figure 5.13: Scheme of home- and work-related tours....96
Figure 5.14: EOD travel times by modes and relation to secondary activity locations....99
Figure 6.1: Schematic search space ellipse....106
Figure 6.2: Search space ellipses – spatial constraint....107
Figure 6.3: Search space ellipses – time and spatial constraints....108
Figure 6.4: Search space ellipses – land-use attractiveness....111
Figure 6.5: Search space ellipses – final choice sets....114

Figure 6.6: Secondary location choice, comparison of EOD and model-1 by mode 122
Figure 6.7: Secondary location choice, comparison of EOD and model-2 by mode 126
Figure 6.8: Visualization of spatial path flows 127
Figure 6.9: Work starting times and durations – EOD and model 128
Figure 6.10: Daily travel time: EOD and model 129

Annex-Figure 1: Duration of primary activity by city sector and user groups 152
Annex-Figure 2: Daily travel times by mode combinations – model and EOD 153

List of Formulas

Formula 1: Calculation of ellipse vertical axis..106
Formula 2: Local attraction factor..111
Formula 3: Calculation of the spatial path flow, example TpTpTp...118
Formula 4: Error E_m between model and EOD 121

Zusammenfassung

In der vorliegenden Arbeit wird ein tour-basierter Ansatz zur Modellierung von Aktivitätenketten unter Berücksichtigung von raum-zeitlichen *constraints* entwickelt. Dazu werden abhängige Wahrscheinlichkeiten für Aktivitäten(-zeiten), räumliche Zielgelegenheiten und Verkehrsmittel berechnet. Dabei erweitert der Ansatz die wegebasierten Verkehrsmodellen zu Grunde liegenden Charakteristika um Aspekte, die in der Regel in aktivitätenbasierten Modellen Anwendung finden. Den theoretischen Hintergrund der Arbeit bildet ein hierarchisch organisierter Entscheidungsprozess, der auf einer Reihe von grundlegenden Regeln entwickelt wird. Diese Regeln werden in eine Abfolge überführt, die das Erstellen eines Berechungsalgorithmus ermöglicht. Das zentrale Argument ist, dass empirische Grenzwerte für raum-zeitliche *constraints* identifiziert und genutzt werden können, um die theoretisch möglichen Entscheidungskombinationen zu reduzieren und damit eine wahrscheinlichkeitsbasierte Berechnung erst zu ermöglichen. Die Ausführungen zur Empirie umfassen die Beschreibung von Aufbereitung und Analyse der Mobilitätsbefragung Santiagos sowie statistische Analysen auf Basis der Beispieltour ‚Wohnen-Arbeit-Sekundäraktivität-Wohnen'. Diese Aktivitätenkette dient auch als Beispiel für die Berechnung der Verkehrsnachfrage. Unter Verwendung eines GIS werden so genannte ‚Suchräume' (Aktionsräume in denen potenziell Sekundäraktivitäten durchgeführt werden) ermittelt. Die Berechnung der Wahrscheinlichkeiten für die raum-zeitlichen Pfade erfolgt über die Erstellung einer Programmsyntax in SPSS.

Ein Ergebnis der Datenanalyse sind Grenzwerte der maximalen täglichen Reisezeit für eine Reihe von Modus-Kombinationen. Die Zeitfenster von Startzeiten und Aktivitätendauer werden in Abhängigkeit sozioökonomischer Gruppen ermittelt. Die Bestimmung der Suchräume erfolgt in Abhängigkeit der Arbeitsdauer sowie der Distanz zwischen Wohn- und Arbeitsort. Beide Kriterien erwiesen sich in der Analyse als statistisch signifikant. Die täglichen Reisezeiten sowie die Suchräume beschreiben die raum-zeitlichen *constraints* des Modells. Der Vergleich zwischen Modell und Empirie (Santiagos Mobilitätsbefragung) deutet darauf hin, dass die Suchräume geeignet sind und die Mehrheit der beobachteten Zielwahlentscheidungen beinhalten. Zur Berechnung der Wahrscheinlichkeitspfade von Aktivitäten und Wegen, wird ein im Umfang auf sieben Ziele reduziertes Alternativenset pro Wohn- und Arbeitsstandort bestimmt. Dabei werden Erreichbarkeit und Attraktivität der Ziele innerhalb des Suchraumes berücksichtigt. Die erzielten Ergebnisse stützen das Argument, dass die raum-zeitlichen *constraints* eine effektive Reduktion der kombinatorischen Vielfalt zulassen. Die Erfahrungen aus der Berechnung der Beispieltour eignen sich zum Übertrag auf weitere Tour-Typen, um eine Modellierung der städtischen Gesamtverkehrsnachfrage zu ermöglichen.

Schlagwörter: raum-zeitliche *constraints*, Wegekettenmodell, hierarchischer Entscheidungsprozess, städtischer Verkehr, Santiago de Chile

Abstract

In this thesis we develop a tour-based approach for modeling activity and travel pattern considering time-space constraints. The estimation of tours is realized through the calculation of interrelated probabilities for activities, locations, modes and times. Thereby, the approach extends characteristics of trip-based transport demand models by including aspects typically treated in activity-based models. A hierarchical structure of choice-making builds theoretical background for the model and is based on a set of axiomatic rules. We transfer these rules into a step-by-step approach resulting in an operational calculation procedure. Our central argument is that empirical thresholds for time-space constraints can be identified and used for reducing the number of choices and, respectively, control the combinatorics associated with the probabilistic approach. In the empirical analysis, results of data mining of Santiago's travel survey are described, followed by statistical analysis of an exemplary tour of type 'Home-Work-SecondaryActivity-Home'. This tour reflects the use case of the approach. In addition, we estimate the so-called search spaces (potential areas where secondary activities are realized) applying GIS. The calculation of the tour path probabilities, that is, the implementation of the step-by-step methodology, is realized using SPSS programming syntax.

From the data analysis we identify thresholds for the tour-based maximum daily travel times considering a set of mode combinations. We define regimes of starting times and duration of activities depending on socio-economic user groups. The estimation of search spaces is realized considering the time spent at work as well as the distance between the home and work locations. Both criteria were found to be statistically significant. Search spaces and daily travel times represent the time-space constraints. The comparison of modeled results with survey observations allowed concluding that the search spaces are realistic since they capture most of the observed trip destinations. For the estimation of spatial path flows, i.e. of the probability of traveling the Home-Work relation, to any secondary activity and back home, we define a final choice set of no more than seven alternatives per primary location considering zone-based accessibility and land-use attractiveness. The obtained results support the argument that time-space constraints allow an effective control of combinatorial complexity. Basing on the experience obtained in process of modeling the exemplary tour, the approach can be applied to further tour types offering the possibility to estimate the entire transport demand of Santiago city.

Key words: time-space constraints, tour-based model, hierarchical choice-making, urban transport, Santiago de Chile

1 Introduction

Decision-making in transport is often argued to be a complex process. There are situations of individual decision-making, such as questions whether to take a car to go to work or to use a bicycle instead. Other decisions occur on a larger scale, beyond the individual perspective and deal with issues like where future roads should be built, if fees shall be charged or if underground lines have to be constructed. In transport, the complexity seems to appear when the distinct levels of decision-making start to interact. Noticeably, what has been described so far is the interaction between transport demand and supply, in other words, between infrastructure and networks, i.e. physical layer on the one hand, and an individual and his or her decisions, preferences and values, on the other hand. Complexity emerges when these levels need to be adjusted and their mutual dependency has to be considered. The decision to provide a certain type of physical transport layer will influence the respective decisions of individuals at present and in the future. At the same time, an individual changes his behavior and his demand for the type of transport supply is changing over a lifetime. A family is founded, a car is purchased, jobs or residence locations change – in each of these cases an individual may expect a qualified physical layer, i.e. transport supply to fulfill the changing needs.

From the first paragraph one can conclude that there is a need to investigate the interplay of transport demand and supply and at best to provide methods or models to evaluate and quantify changes in the system, e.g. due to policy interventions. This objective is not new: research on the development of transport models aiming at the reproduction of demand and supply has been carried out for almost 60 years. But in the end, all these models attempt to disenchant the complexity of demand and supply and to make cause-effect relationships more explicit. What happens if we build a road or an underground line? How does rising car ownership influence daily travel behavior? As models have been gaining in details, the questions that researchers aim at answering have been changing over time, too. The field of models and methods applied is rather broad; nonetheless, it is possible to detect underlying ‘main characteristics’ or ‘methodological streams’, and one of the opening chapters will deal with this issue.

The present thesis contributes to the developing of methodologies for the modeling of demand-supply interaction. The proposed method integrates aspects of different ‘methodological streams’, namely those of the established aggregated approaches (four-step models) and the newer activity-based models. In the thesis we intend to ‘walk all the way’ from setting up a methodology of how transport demand can be interpreted, towards data analysis and up to testing the model through its empirical application. Respectively, the target is to render a methodological contribution that would reduce the gap between aggregated and disaggregated models.

1.1 Objectives and approach

This work is motivated by the interest in better understanding of individual travel behavior. Besides gaining this profound understanding, the objective lies in achieving the transfer of what is roughly defined here as 'travel behavior' into an empirical application. This objective is pursued in three major steps: first, development of a methodology of how travel behavior is interpreted; second, analysis of travel behavioral data; third, setting up of a calculation procedure (also called model, approach or empirical application). Each step is accomplished considering the central research question: what time and space constraints on travel-related decisions are appropriate to be considered and how should values for them look like to enable the modeling of activity and travel patterns?

The challenge is to provide and to test a methodology of generic validity, which brings up the task of developing a transferable approach. It means that the underlying methodological principals should be adaptable independently of the spatial context. Naturally, a certain amount of data needs to be assured. However, given the data available, the approach has to be applicable in different spatial contexts. Although this characteristic should be the basic principle of any method provided, it is worth mentioning that the 'Santiago case' and the respective information sources used represent only one specific application of a more generic approach.

Another objective of this study deals with one of the well-known challenges in transport demand modeling: definition of a reasonable trade-off between the model's behavioral details considered and the simplifications applied to maintain the complexity manageable. It was mentioned that in general terms two large 'methodological streams' are observed distinguishing between an aggregate interpretation of transport demand, where probability flows of socio-economic clusters are calculated (trip- or chain- based flow models), and disaggregated approaches, where individuals are simulated along each step of their travel decision-making (activity-based simulation models). We pursue the objective to estimate activity and travel probability flows by household clusters. However, this perspective is expanded because we interpret demand as an interrelated sequence of activities and trips, i.e. tours and/or activity patterns are subject of this research. For instance, choice of location for an activity depends on the previous location, on the time spent there or on the remaining travel time available and on the transport modes used.

Among many other questions and issues treated in models of transport demand, particularly the solution of the destination choice problem is crucial to any approach. Destination choice is often treated by estimating the probability flows between origins and destinations, by simulating individual choices assuming rational utility-maximizing behavior or by applying heuristic rules. However, the solution of the location choice turns to be a true challenge, since the objective is to estimate the expected travel flows using e.g. probabilities as one method but at the same time to comply with the criteria of choice dependencies. This can easily be illustrated if we imagine that a person does some private errands after work. Theoretically the entire area, e.g. a city, may provide opportunities to fulfill this need. If we then add different modes to this choice situation and different time periods in which the private errand might be run, a huge number of rather small probabilities for each choice combination appear. This complexity, expressed by the number of combinatory options, becomes an exponential problem if we assume further choice dimensions.

In consequence to the previous paragraph, the thesis attempts to provide an approach that would reduce the huge number of choice combinations. It seems rational to assume that not all actually feasible choice options should be taken into consideration at any point in time. If individuals take time to evaluate all choice options of e.g. mode and destination combinations, the society most probably will get stuck as the 'evaluation of options' would consume the entire resources, i.e. all the time available. However, a model, which considers both the likeliness (probability) of different options and still accounts for the dependencies between choices, requires adapting characteristics of different models (and methods) that already exist. In this sense, we present a sort of a 'hybrid' approach applying characteristics of both aggregated and disaggregated methods. The method constitutes on a set of main principles which are followed throughout the analysis and application. Because of their relevance, we already introduce them as follows:

- **Scales:** We assume that decisions have a graduated range since individuals assign priorities to their activities. The scales are independent and conditional. For instance, higher priority for activities, locations or modes defines subsequent choices.
- **Time and space:** It is observed that travel behavior is guided by the constraint of time. For each individual this resource is limited, and so is the reachable geographical space accessible given the time available.
- **Cognitive restriction:** A person has a naturally limited ability to process and evaluate all the available choice options. It seems to be a rational and behavioral sound assumption that only a subset of choice options is taken into consideration out of the actually feasible number of options.

Based on these principals a method for modeling transport demand is developed. We strictly follow the principal of 'scales' assuming that decisions related to a subordinated activity of e.g. shopping or leisure are adjusted according to superior decisions related to e.g. the work activity. The constraint of time implies a limited geographical space where activities may be realized. Finally, we use the argument of 'cognitive restriction' to further reduce the number of choice options to a behaviorally sound set used for the estimation of the demand.

In practical terms we develop a statistic model aiming at the reproduction of empirical observations. The statistic model considers time-space constraints and is based on a step-by-step methodology to model hierarchical choices in urban transport demand. A comprehensive travel survey, Santiago's 'Encuesta Origen-Destino, EOD', is used for the estimation of empirical probabilities of activity and travel behavior. In addition, output of the city's land-use and transport models is considered, for instance, regarding level-of-service by mode, commuting matrices or Santiago's land-use. The EOD is subject of data mining procedures aiming at identifying the coherent patterns of activities and trips. This is followed by an empirical analysis deploying descriptive statistics, regressions and correlation analysis for the identification of patterns and relationships between the data. With the objective of defining delimited areas where activities are conducted, ellipses are estimated making use of a Geographical Information System (GIS). Eventually, all empirical parameters enter into a computational calculation algorithm, used for the estimation of constrained spatial path probabilities.

1.2 Chapter structure

The thesis is divided into two major parts; the first one presents the theory of travel behavior decision-making and introduces in a generic way the developed method for the modeling of activity and travel pattern. The second part deals with the empirical analysis and the implementation of a calculation algorithm, its application and evaluation. Overall, the thesis is structured into seven chapters as presented in Figure 1.1.

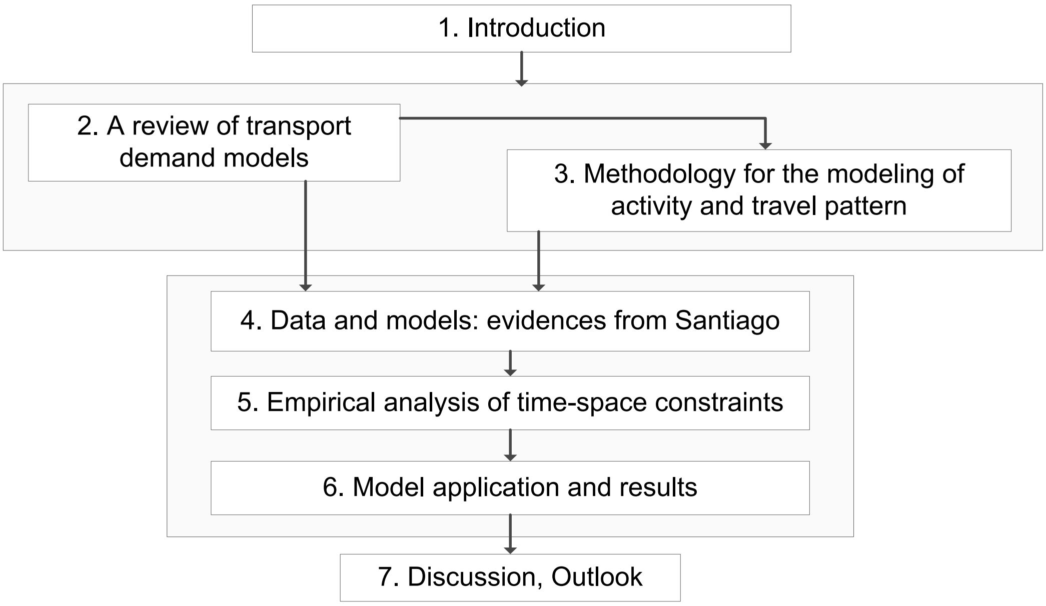


Figure 1.1: Chapter structure

Chapter 2 and chapter 3 represent the bibliographic review and the theory formulation. Both provide input to the upcoming chapters defining the underlying methodology of the developed approach. This has implications on the strategy of the empirical analysis (chapter 4 and 5) as well as on building of the model application in chapter 6. Chapter 4 summarizes the main data mining steps conducted on the EOD. Chapter 5 illustrates the statistical analysis of time-space constraints with the purpose of quantifying relevant model parameters for the empirical application. Chapter 6 reveals the major results and discusses the findings from the application. Additionally, chapter 6 focuses on the validation of the results against empirical observations. The final chapter 7 discusses the results and the approach as a whole, focusing on identified strengths and weaknesses. Chapter 7 closes with an outlook on further research needed to continue refining the established methodology.

2 A review of transport demand models

Any travel demand model represents an attempt to reproduce travel behavioral choices in a given spatial context. Numerous methodologies and models have been developed so far and some of them are commercially distributed as standard software packages. Models vary significantly due to different interpretations of time, space, socio-economic attributes of the population and due to different modeling of choices for transport modes and locations. Respectively, different categorizations are possible when providing a review of their main characteristics. Basically, the following categorization refers to the one suggested by Hensher and Button (2008), and by Cascetta (2009). The categories are shaped by the manner in which transport demand, i.e. the number of activities and trips, is treated and therefore fall into three major categories: trip-based models, trip-chaining models and activity-based models. To some extent, the order in which the model types are presented reflects their appearance over time, the trip-based models being the earliest applications and the activity-based models representing the latest research. Another approach would distinguish between the decision-making theory and models used e.g. for mode and destination choice. However, these theories and methods represent an indispensable part of the major categories mentioned above and are discussed within the respective sections. It will be shown that to a certain extent similar methods are applied throughout the categories.

We focus on the underlying philosophies of how activity participation and travel demand are actually treated. The task is to identify general characteristics of each approach and to discuss their respective strengths and weaknesses, which is relevant for correlating the modeling approach developed in this thesis to the body of earlier research. We think it is important to devote one section to the relation between time geography and models of transport demand. The fundamentals of time-space constraints represent a reference for almost any transport model – independent of the belonging to one of the categories we describe – and are of major relevance in this work, too. The summary section discusses briefly what elements of the revised models are planned to be considered in the approach to develop. It is important to point out that the detailed mathematical foundations of each approach are not provided save for the respective references.

2.1 Trip-based approaches

The fundamental unit of the analysis of trip-based approaches is an individual trip. Each trip between origin and destination is modeled independently of other choices, i.e. the interrelationships between trips and activities are ignored. This characteristic remains one of the main limitations of the trip-based approach, the fact that has set off the research on the behaviorally sounder approaches (McNally and Rindt, 2008, p. 57). A trip-based model typically involves four steps of travel demand modeling. These steps comprise trip generation (1) when the average number of trips is being basically estimated by purpose, by user group and, eventually, by time period, too. The number of trips is often directly estimated via a classification table (trip rate by user group and purpose) with trip frequencies indicated basing on the analysis of travel survey data. To this end, the sequences of reported activities in travel surveys can be split into single trips, thus assigning trip frequencies to specific user groups by trip purpose (Lohse, 1997, p. 19).

A trip-rate regression model represents one of the more sophisticated solutions and interprets trip rate as a linear function of socio-economic variables and car ownership. Another approach is the application of a behavioral random utility model where the trip rate again is defined as a dependent categorical variable while socioeconomic attributes and zonal attractiveness reflect the independent variables. In this case, choice alternatives for the dependent variable representing different numbers of trips undertaken are included. As a result, the utility (or probability) is estimated applying e.g. a multinomial logit model. Given the socioeconomic variables and the spatial characteristics, the utilities for conducting 0, 1, 2 or more trips are estimated within this model (Cascetta, 2009, pp. 181-5). Independently of the specific method applied, the results of step (1) are trip frequency rates, which, if multiplied by the population data set – generally a segmented population in behavioral homogeneous groups – represent the overall number of trips for the study area.

The second (2) and third (3) steps of the four-step model refer to the destination and mode choice respectively. The two steps are named here concurrently, as they are sometimes modeled simultaneously as well. This is meaningful to assure that persons only travel to those destinations they can access given e.g. the car availability or the general accessibility of the specific zone by transport modes. However, other types exist treating mode and destination choice separately forming the following two model types (Friedrich and Schiller, 2010, p. 113):

Trip-end models: they imply that mode choice is estimated directly after the trip generation; respectively, different zone attractiveness by mode is not considered; in fact, it is assumed that personal characteristics, in particular availability of a car, determine adjacent choices; this assumption makes trip-end models primarily useful for the estimation of car travel demand, neglecting the influence of different public transport qualities on the decision-making.

Trip-interchange models: they imply that the destination choice is estimated first, whereas mode specific qualities, i.e. different travel times can be considered as well; this is a major difference to trip-end models where actually no choice between different modes based on their level-of-service attributes is realized; if the destination choice is made in a more disaggregated fashion by user groups and by trip purposes, personal characteristics can also be considered.

Within the destination choice model the number of trips is assigned to the particular destination zones by origin, purpose and, eventually, by time period. Within trip-based models a gravity model is frequently applied, where the number of trips between origins and destinations is proportional to the attractiveness between zones. Each destination has a specific attraction potential, for instance, reflected by jobs, schools, retail or services space. The classical gravity model considers only the total amount of trips and distributes them between origins and destinations interpreting the potentials of production (origin) and attraction (destination) as weights (Cascetta, 2009, pp. 185-6; Friedrich and Schiller, 2010, pp. 100-1). In addition, the production or attraction side (or both) can be handled in a constrained manner, resulting in either a single or a doubly-constrained gravity model (Bates, 2008, p. 27). In the former case either the number of trips of the production or of the attraction side is fixed, in the latter case both marginal totals need to be met. The availability of marginal totals can be based on specific knowledge, e.g. on observations or counts either on the production or on the attraction side. In case of the zone attraction, it is predefined that some destination zones attract at maximum a specific amount of trips. This is especially meaningful for work or study trips, where the number of jobs or school places indicates the capacity limit. In cases of other travel purposes such as shopping or leisure trips the definition of a capacity constraint is less straightforward as it is difficult to define maximum capacity limits precisely. In practice, all types of single and doubly-constrained gravity models are applied. In the meantime and in addition to the fixed marginal totals, so-called elastic capacities were introduced in some models, where capacity ranges of minimum and maximum marginal totals are considered for discretionary travel purposes (Schiller, 2007, pp. 56-7).

The gravity models discussed so far have not yet considered the behavioral aspects, for instance, the difference in perception of the attractiveness of OD relations by various user groups. For this purpose, similar to the trip frequency models, random utility models in form of a multinomial logit model may be applied. These models allow for the estimation of behavioral parameters that represent the individual or user group based evaluation of OD relations. In practice, parameters reflect different perceptions of similar decision-making situations: for instance, the travel time spent on a leisure trip is perceived as ‘less negative’ compared to the time spent commuting. The estimation of the respective parameters requires the definition of a choice set of destination alternatives and claims the respective level-of-service attributes. Models consider variables that describe simultaneously the destination attractiveness and the specific attributes of the related mode choice. Thus, beside zonal attributes such as size of a facility, various cost attributes such as walking time, in-vehicle time, monetary cost, etc. may be included. The resulting parameters are used to estimate the probability that an individual (eventually pertaining to a socio-economic user group) chooses a mode-destination combination (given the variables represented in the model, i.e. the choice situation). In trip-based models, the parameters may be applied during the solution of the simultaneous step of mode and destination choice. Simultaneous models estimate the probability of traveling between all ODs considering a modal split (by user group and by travel purpose), the impedances between all ODs by mode and travel purpose as well as the parameters that reflect different perceptions of travel time savings (losses) among travel purposes. Further attributes describing the transport mode quality such as comfort, reliability or security may be considered as well. However, in trip-based approaches as well as in other models the quality attributes are often reduced to time and cost, since further information is rarely available.

From a behavioral perspective it makes sense to integrate mode and destination choice since the respective decisions are not taken independently. The objective and result of any trip-based model type are origin-destination matrices differentiated by mode. In the final step, i.e. in the route assignment (4), the matrices are assigned to network models of public and private transport. In case of car travel the link-based and the route-based demand flows lead to alterations in the level-of-service attributes (e.g. due to congestion) that in turn have an impact on the models of destination and mode choice. This is important since level-of-service has an influence on the attractiveness of the zone, thus on the probability of being chosen as an option for destination choice. Iterations between demand and route assignment are repeated until no further change in route choice and, subsequently, in mode and destination choices is observed. This situation plays the role of a user equilibrium and embodies an important feature of many four-step models, since the equilibrium offers the final solution as no further alterations between transport demand and supply occur. The following generic Figure 2.1 summarizes the elements of the four-step model and illustrates the interaction between transport demand and supply.

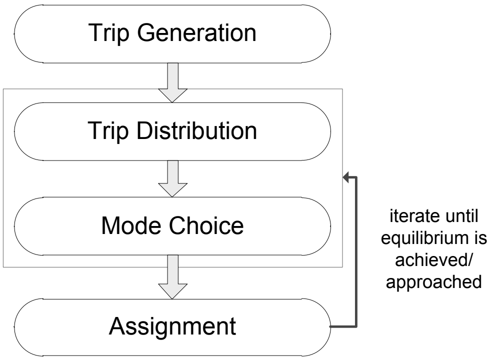


Figure 2.1: Scheme of the four-step algorithm

The description of trip-based approaches provided here remains rather generic without an in-depth review of functionalities or mathematical foundations. However, it seems sufficient at this point to focus on the general characteristics and relate them to those incorporated in the approach pursued in the thesis. Previous research has demonstrated some strengths and certain limitations associated with trip-based models. The following characteristics can be summarized regarding models' strengths:

- The application of gravity models including different underlying functions for the evaluation of alternatives represents a robust methodology.
- The application is possible with relatively low data requirements due to generally few user groups and travel purposes.
- The trip-based models allow achieving equilibrium between transport demand and supply, actually representing a unique solution to a complex problem.
- Procedures for the calibration of model parameters, especially for destination choice, are established and are often included in software packages.

Nevertheless, many authors argue that trip-based models lack of a sound behavioral background. Further research on other model types, such as trip-chaining or activity-based models, has been undertaken. In sum, the following limitations of trip-based models can be listed:

- The behavioral relationship between activities and trips, thus, the conditionality between choices, is neglected.
- The temporal dimension, i.e. the time trips start, is often introduced afterwards only as a post-processing step using starting time distributions observed in travel surveys.
- Trip-based approaches are mostly based on individual user groups; thus, effects of the household context on travel-related decisions (e.g. activity participation, scheduling or joint car-use) are not considered.

Trip-based approaches continue being the most frequently applied travel demand models in practice today. Although some newer approaches are already available, they have not gained the popularity of trip-based models yet. However, the advantage of relatively low data requirements and the ability of a unique solution (equilibrium between demand and supply) are important aspects speaking in their favor. Nonetheless, not considering dependencies between choices appears to be the most considerable weakness from a behavioral point of view.

2.2 Trip-chaining approaches

Trip-chaining approaches explicitly address the aspect of conditionality between choices. Travel demand is no longer interpreted as a sum of movements between origins and destinations, but as an interdependent sequence of activities and trips. Thus, in the phase of trip production, number and types of activities and trip sequences determine the overall number of trips. To this end, a set of trip-chains and probabilities of being conducted by user groups are defined (Lohse, 1997, pp. 20-2). This denotes an important difference from trip-based approaches, since the observed activities and trip sequences of the survey are not divided into single trips but are processed further as an entire chain. This perspective appears to reflect more closely the way in which travel decisions are actually made (Frank et al., 2008, p. 39). The main difference to the activity-based approaches is that trip-chain models as interpreted here are normally based on user groups, aiming at the estimation of chain flow probabilities, while activity-based approaches simulate individual decision-making. A trip-chain typically starts and ends at home and is represented by one or more activities and two or more trips, respectively. The term trip-chain may be analogous to the term ‘tour’. Especially in North America, transport models based on the processing of interdependent activities and trips are defined as tour-based models. Respectively, a tour is also defined as a sequence of trips starting and ending at the same location with one or more intermediate stops (Axhausen, 2008, p. 332; Golob and Hensher, 2007, p. 299). However, an important distinction has to be pointed out here: the characteristics of trip-chains or tour-based models are discussed in this section assuming that a fixed trip-chain is an input for demand generation. This is important as in the section devoted to activity-based approaches, the term ‘tour’ is used as well, but it implies that activity scheduling, i.e. building of an activity sequence, forms part of the model.

For the sake of providing a clear handling of terms in this work, we will provide an unambiguous terminology in one of the upcoming sections.

Further definition of the concept of tour-based or trip-chain approach is given as follows: "These models group trips into tours based on the fact that all travel can be viewed in terms of round-trip journeys based at the home. A tour is assumed to have a primary activity and destination that is the major motivation for the journey" (Bowman and Ben-Akiva, 2000, p. 3). Tour-based models do consider the request to interpret transport demand as a set of interdependent choices; nevertheless they fail to reproduce the entire daily travel behavior that sometimes consists of a number of tours. Bowman and Ben-Akiva summarize this as follows: "However, the tour-based approach lacks a connection among multiple tours taken in the same day, thereby failing to capture the effects of inter-tour temporal-spatial constraints (ibid, p. 3)." Consequently, a set of tours conducted on one day, defined as an activity pattern, has to be considered in addition to the concept of tours or trip-chains. However, the current section focuses on single trip-chains or simple tour-based approaches. Respectively, models of this type estimate tours, but not the entire activity pattern. Hence, some information is lost since interdependencies between various tours made by the same person are not considered. Concepts of simple and complex tours are exemplified in Figure 2.2, whereas an activity pattern may distinguish between primary activity (work taken as an example here) and any other activity.

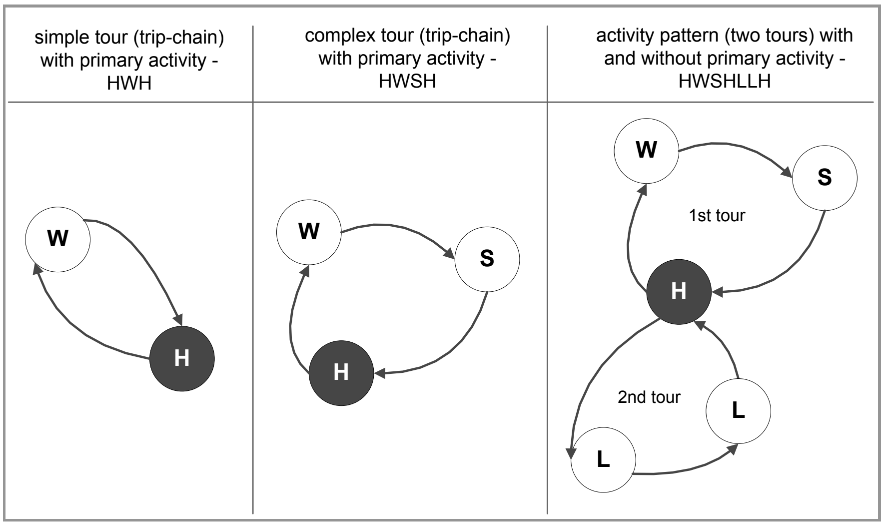


Figure 2.2: Concepts of tours (trip-chains) and activity pattern

Annotation: H = Home; W = Work; S = Shopping; L = Leisure.

There is a difference not only between tours and activity pattern but also between activities, depending on their priority. Trip-chaining approaches often classify activities (travel purposes) into those of primary and of secondary importance. Most generally, primary activities are those of compulsive character, such as work, school or study, while secondary activities are those of discretionary character, such as leisure or shopping. Regarding the mathematical foundations of these models, there is no single characteristic mathematical approach. The applied methods are rather similar to those discussed in the context of trip-based models, such as random utility maximization theory for e.g. the mode choice (Miller et al., 2005, p. 401) or gravity models for destination choice (Fellendorf et al., 1997, p. 61). Differently from the trip-based approach, the mode choice alternatives can also be represented by mode combinations. For this purpose it is considered to reduce the range of theoretically feasible mode combinations. For instance, it is assumed that if a car or a bike is used for the first trip, the same mode is used for subsequent trips along the tour. In the case of walking or public transport trips, freedom of choice is given regarding the mode alternatives chosen along the tour. However, similarities to trip-based approaches appear in the step of trip production. Travel survey data provide a fixed number of trip-chains and, thus, the number and sequence of activities and related trips. The estimation of the number of trips by travel purpose is then the product of the number of persons (households) of any origin and the fraction of a respective trip-chain. For instance, 100 persons follow the trip-chain HWSH (Home-Work-Shopping-Home) and provided this trip-chain appears with 5% probability, it totals in 15 trips (100 * 5% * 3).

There are only few trip-chaining approaches actually applied in practice. As argued so far, in a certain way they represent an intermediate solution between trip-based and activity-based approaches. In the past decades much more attention was paid to the development of activity-based approaches. Common between them is that they interpret transport demand as derived from interrelated activities, while the difference is that in case of activity-based approaches the subject of analysis are individuals and detailed decision-making situations rather than prediction of aggregate probability flows. The reason for a lack of probabilistic tour-based approaches might be the increased complexity of the task to estimate interdependencies of decisions. The more variables are involved (e.g. time periods, socio-economical factors, modes, etc.) or the more complex the tour or activity pattern to model gets, the more difficult it is to handle the potential combinatorial options. Respectively, in cases where many variables are considered it gets likely that decision-making is simulated individual by individual rather than predicted based on conditional probabilities. This observation is of relevance, since the objective pursued in this work is to handle complex decision-making, thus complexity and combinatorial issues, without departing from the concept of conditionality between choices.

Further on, one example of a trip-chaining approach will be explained in more detail. The dependencies between decisions will be dealt more explicitly along with identification of the approach's strengths and weaknesses. Program system VISEM, described in Fellendorf et al., 1997, serves as an example.

VISEM: a program system to estimate transport demand

The underlying methodology of program system VISEM can be defined as an activity chain based model, primarily developed for applications in urban areas. As described above, trip frequencies are estimated directly basing on travel survey data about trip-chains and referring to order and sequence of activities and trips in the course of one day. Trip-chain types observed in travel surveys vary significantly: typically, several hundreds of different types can be detected. Respectively, VISEM considers only a limited number of trip-chains. The population is grouped into behaviorally homogeneous subgroups of persons, differentiated by status (employed, unemployed, students, schoolchildren, young children) and by car availability. Probabilities of trip-chain types chosen in each group are derived directly from the data.

In most of the applications, VISEM has been applied for estimating trip-chains of up to three consecutive trips, i.e. trip-chains of HWSH (Home-Work-Shopping-Home) type. Decisions on destinations and modes are partially modeled in an integrative manner. The choice of the first destination (here the work activity) is estimated using a distributive function with parameters specified for the combination of person group, activity (travel purpose) and mode. These parameters are derived applying a weighted linear regression based on empirically observed travel distance distributions (Fellendorf et al., 1997, p. 63). As level-of-service attributes are considered in the regression, mode characteristics influence the destination choice. The attractiveness of zones depends on their land-use properties, e.g. number of jobs, retail floor space or capacities of leisure facilities. Given the use of a distributive function specified for the location choice, even remote locations of less attractiveness obtain at least a minor probability within the matrix. This implies that the number of feasible choice options is high and so is the number of choice combinations once a trip-chain consists of more than one single activity. Against a detailed zoning system the issue of small probabilities for too many choice combinations can become a computational challenge and it is debatable whether these choices with very low probabilities represent reasonably behavioral decision-making.

Within VISEM the probabilities of mode choice are estimated for each group and for each trip purpose applying a multinomial logit model. The model considers the socio-economic situation of each personal group, i.e. car availability, the attributes of each mode and predefined choice restriction regarding the combination of modes. Parameters of the utility function comprise travel time, time for access and egress, distance, and fare as well as the constant utility by mode. Mode choice is estimated for the first trip and if this is an exchangeable mode (public transport, walking), a second, independent mode choice is realized for the second trip. Otherwise, the first mode choice determines subsequent mode choices, i.e. stays fixed for the second trip (if a car or a bicycle is used). Finally, travel demand, i.e. number of trips from origins to destinations considered by modes, is estimated as a product of the number of persons by group, the trip-chain probability and the combined probabilities of modes and destinations.

Temporal dimensions of activities and trips are not considered within VISEM. Once trip tables are estimated by purpose, starting time distributions for specific activity combinations (for example, Home-Work) are used to timely disaggregate the demand. Empirical travel survey data serve as a source for these distributions. This means that, for instance, time spent on an activity does not affect the choices related to the subsequent activity. Neither the total travel time needed for the entire trip-chain influences its overall feasibility (probability).

Basing on general descriptions and on the example of VISEM, the following strengths of trip-chaining approaches can be summarized:

- The modeling of dependencies between activities and trips is explicitly addressed, and conditionality of choices is considered.
- The iterative consideration of updated level-of-services is possible; thus, equilibrium between demand and supply can be achieved.

- Data necessary for trip production and calibration of mode and destination choice is derived from regular travel survey data; thus, data requirements are comparably low to those of trip-based approaches.

Though the issue of interdependency between choices is dealt with within trip-chaining approaches, the following shortcomings can be revealed:

- The modeling of dependencies is limited to single tours. Effects of inter-tour relationships are not considered.
- The probabilities for a trip-chain (including decisions on locations and modes) are estimated without taking into account temporal or spatial constraints, e.g. a limited daily travel time.
- The appliance of gravity models for destination choice implies a serious problem with regard to combinatorial numbers. Even remote destination zones (disregarding any criteria of temporal or spatial constraint) showing a negligible probability (feasibility) remain part of the feasible choices.
- Similar to trip-based approaches, the temporal dimension, i.e. assignment of trip starting times, is introduced only after the decisions on modes and destinations have already been modeled.

The most important contribution of the trip-chaining approach is that trips are modeled in an interrelated way. This represents a crucial step forward to a behaviorally sounder framework of transport demand models. However, strong criticism remains. Logically, the ongoing research seeks to overcome the shortcomings. The criticism triggered what was called a ‘paradigm change’, moving research from the perspective of aggregated trip flows realized by behaviorally homogeneous groups towards the perspective of an individual and his decisions on activities and trips.

2.3 Activity-based approaches

According to the underlying model philosophies, the appearance of activity-based approaches means a remarkable change in how activity and travel behavior is understood and modeled. The fundamental principle of activity-based approaches is that travel is interpreted as a demand derived from the necessity to conduct activities. In difference to trip-based approaches, activity-based approaches focus on the activities that prompt trips (McNally and Rindt, 2008, p. 59). This fundamental change in the perspective is the main characteristic of this approach and is confirmed by different authors. Builing and Kanaroglou (2007, p. 151) state that, “underlying the approach is the widely accepted view that travel demand emerges in response to individual and household requirements for activity participation.” Similarly, Davidson et al. (2007, p. 465) conclude that “an activity platform implies that modeled travel is derived within a general framework of the daily activities undertaken by households and persons.” Bowman and Ben-Akiva (2000, p. 2) bring it to the point resuming that “the demand for travel is derived from the demand for activities.” Besides focusing on activities rather than trips, they name another fundamental characteristic of activity-based approaches:

"humans face temporal-spatial constraints, functioning in different locations at different points in time by experiencing the time and cost of movement between the locations (ibid.)." Especially the latter point is of relevance as it brings up the point of temporal and spatial constraints. In consequence, activity-based approaches generally seek to consider those aspects in the modeling approaches. The constraints themselves can be manifold, ranging from opening hours of facilities to the necessity to coordinate activities with household members.

Generally, activity-based approaches are applied in the context of urban agglomerations. In this context, Cascetta says that "(…) the complexity of urban living has resulted in correspondingly complex trip-making behavior" (2009, p. 228). This complexity – if true – is often reflected by activity-based approaches that include much more variables than the previously described approaches. Some authors relate the emergence of activity-based approaches also to the need of overcoming limitations associated to trip-based models that were not able to evaluate policies of transport demand management appropriately. The shift in the perspective from provision and evaluation of the infrastructure towards the analysis of e.g. regulatory measures (e.g. pricing schemes, parking restrictions or new mobility options) meant new challenges to the established models, and they were not able to cope with those (TRB Special Report 288, 2007, p. 101). Another important characteristic of the newer models is that they normally work at a disaggregated level, say, at a level of individuals or households. However, the methods and approaches applied for modeling individual choices are partly the same as in the previously described approaches partly differ from those to some extent. Random utility models are applied for estimation of individual preferences in choices as well. In cases of (microscopic) simulations individuals and their decisions are simulated one-by-one considering elements of (weighted) random behavior. Other approaches are rather based on rules, with agents (individuals) applying behavioral heuristics derived from survey data or personal interviews (Timmermans et al., 2002, p. 184). These decision-making rules then serve as an equivalent to the parameters estimated in models based on utility-maximization.

Other types of rules considered in activity-based approaches reflect dependencies between individuals taking joint decisions in the household context. For instance, if a person has taken the car in the morning, the car as a mode option is not available anymore for other members of the household. Or, bringing a child to the kindergarten implies the constraint for somebody of the household as he or she has to pass by at a certain time of the day to pick up the child (Petersen and Vovsha, 2006; Bradley and Vovsha, 2005, pp. 547-8). From a behavioral point of view the consideration of these dependencies appears indispensable and somehow trivial, as these decisions are part of daily life. But the reproduction of relationships in models is complex and might sometimes result hardly possible due to data limitations. Respectively, the incorporation of household-interactions into travel demand models is still a relatively new body of research. However, these examples illustrate two principal aspects related to activity-based approaches: a) the objective is to offer a sounder behavioral framework and b) scope and detail of the approaches varies a lot resulting in a quite heterogeneous field of different approaches either being under development or already in application.

Especially the latter aspect makes categorization of the models or recognition of their strengths and weaknesses a big challenge. This heterogeneity is an obstacle for the models to actually enter the market and replace trip-based approaches. At present, the actual application of activity-based approaches confines to a few planning entities in the US. The argument that activity-based approaches are more complex and data demanding, whether this is true or not in any case, creates prejudices among users and makes them stick to the well-established trip-based models. However, at the core of the models some main characteristics can be identified, as listed below.

- Travel demand is derived from the demand for activity participation (see discussion above). The unit of analysis is a tour and/or a pattern of activities and trips behavior, not individual trips.
- Often the models attempt to consider that the household context influences activity and travel behavior. In activity-based approaches interpersonal interdependencies are explicitly addressed (for instance regarding car usage, common shopping activities, etc.).
- Activity (re-)scheduling makes it possible to add or to remove activities from an activity schedule depending on the situation (e.g. a skipped business meeting offers time for an additional shopping activity which otherwise would have been postponed).
- The temporal aspect is already incorporated into demand generation, i.e. activities and trips are modeled (simulated) in their timely order.
- Sometimes more detailed spatial systems (subzones, grids) are considered.

Naturally, a model that considers all these aspects is behaviorally superior to trip-based and trip-chaining approaches. Nonetheless, complying with all these criteria leads to a substantially more complex model. In this context McNally and Rindt (2008, p. 61) argue that "while models are abstractions of reality, the reality in this instance is no less than daily human behavior, replete with all its vagaries." A model that fulfills all the named aspects in a satisfactorily manner is definitely more appropriate for the evaluation of (transport) policies. Not only the impact of infrastructure and level-of-services but also, for instance, the effect of a childbirth on the adjustment of activities (and trips) of infant's parents can be measured. To make the previously said more tangible and to better understand how some characteristics of an activity-based perspective are translated into a model, one example is described in more detail.

Activity-based approach

The approach developed by Bowman et al. (Bowman, 1998; Bowman and Ben-Akiva, 2000; Bowman and Bradley, 2008) is denominated as an activity-based disaggregated travel demand model. The main objective is the estimation of an activity pattern, including one or more tours and the respective decisions about activities, modes, locations and the time-of-day these decisions are taken. In the model, an activity pattern comprises one or more tours, defined as a sequence of activities and trips starting and ending at home. In concrete terms, the pattern includes the primary activity of the primary tour and, potentially, further activities, which may lead to more stops (and more trips) within this tour. Additionally, secondary tours can occur after the primary tour, again characterized by the number of stops (and associated trips). Contrary to trip-chaining or single tour-based approaches, linking of tours during the day is an important characteristic. This ensures interdependency of decisions not only inside one tour but in between different tours as well. Thus, the composition (e.g. activity durations) of the primary tour determines the characteristics of the subsequent tour(s) and associated decisions.

Another important characteristic of an activity-based approach is that choices are modeled in accordance with assumptions about hierarchies. The hierarchy among primary and secondary tours has already been mentioned. A priority is assigned to each tour stepwise applying a deterministic approach, where work activity is of highest priority, followed by work-related activities, school and other purposes. If two primary activities appeared within the same tour, the activity with the longest duration received the highest priority. Among tours of a pattern, the priority was assigned to the tour with the activity of highest priority. Thus, an activity pattern consists of one or more tours that are hierarchically ordered and include activities of primary or secondary importance.

Any secondary tour is modeled dependently on the primary tour, whereas no hierarchy is assumed among secondary tours. Thus, the utility of the activity pattern is calculated as the joint probability of primary and secondary tours. The destination and mode choice within each tour is estimated applying a multinomial logit model based on a sample of mode destination alternatives. Initially, the model allowed selecting between six modes, assuming that one mode is used throughout the tour and not taking into account mixed mode combinations. Additionally, the sample was split into four time-of-day categories to consider the effect of different levels of services during the day. In total, the model suggested by Bowman and Ben-Akiva allowed for the selection of 54 different patterns (tour combinations). The probabilities for a single or a multiple set of tours, including decisions about time, mode and destination, were defined for each single person of the respective group of population. As a result, the number of trips by purpose, mode (and time period) could be calculated and assigned to the respective networks through mode-specific OD matrices.

The model developed by Bowman and Ben-Akiva clearly demonstrates some advantages due to the increased level of detail. Their approach addresses a) the interdependency of decisions with the consideration of patterns and tours, b) the temporal dimension of decisions and c) provides the entire set of travel-related information for each individual living in the study area, represented by one agent each in the model. Although the perspective is on modeling individual choices, some aggregations are made for instance through considering time-of-day categories and a subset of activity patterns. In the approach the issue of household interaction via the modeling of jointly realized trips or the arrangement of car use is not included. Even in a version of this model developed later on no models of household interaction were established (Bradley et al., 2010). Nevertheless, household variables such as size, composition and income are considered in the estimation of probabilities for patterns, tours and subsequent decisions.

Another general characteristic of activity-based approaches is that sometimes a more detailed spatial system is considered. Referring to the example above, in the first version of the model the calculation of level-of-service attributes was based on zone-to-zone information, typically provided by network models. Recently, efforts have been made to refine the spatial attractiveness of destinations considering parcel-based information. Still, the estimation of mode and destination choice models is based on a sample set of alternatives, but now based on parcels (Bradley et al., 2010). The approach has been applied in practice in several cities in the US, thus representing one successful example of 'transition' from former trip-based towards activity-based approaches. However, some concerns are still accompanying the development of activity-based approaches. The level of detail considered (time, geography, trip purposes) sometimes creates an unfeasible amount of options, forcing the models to work with population samples. Another concern is the achievement of equilibrium between demand and supply, which stays a challenge due to large running times.

The discussed approach is an example indicating the general objective pursued with activity-based approaches, which is the reproduction of travel behavior in a highly disaggregated manner. This level of detail is achieved at the cost of extensive data needs and running times. It also evident that despite the achieved detail some aspects are treated in an aggregated fashion (e.g. time-of-day, household types, activity patterns).

2.4 Constraints and the role of time geography

The development of trip-chaining and especially activity-based travel demand models is highly motivated by the introduction of more behavioral realism into the models. This was outlined in the previous section about activity-based models. Many of recent approaches in travel demand modeling focus on issues of constraints, conditions or even social interactions. Today's research examines options to introduce even more constraints or conditions, hoping to create more realistic models, and focuses on the individual living in a social context, which is mostly the household. Hence, researchers attempt to incorporate situations of complex decision-making into models, e.g. situations of joint participation in an activity when family members go shopping together or accompany each other.

However, many of these motivations for today's modeling already have a large history of discussion in the scientific area. Torsten Hägerstrand (1970) was the first to provide grounds for claims for more behavioral realism in the reproduction of activity and travel behavior. In his important article "What about people in regional science?" he explains that time and space can be defined as resources that constrain human activity behavior. Although Hägerstrand confirms that there are many more time-space phenomena to consider, he attempts to reduce these to three key categories of constraints. He differentiates between 'capability constraints', i.e. physiological limitations of an individual (the need to rest and to eat); 'coupling constraints', i.e. the need to accomplish certain activities together with other people at specific locations (workplace, school); and corresponding to given restrictions and institutional settings 'authority constraints', i.e. opening hours of facilities and institutions (Hägerstrand, 1970). Until today these constraints represent the basic principles of time geography and pose a major challenge in modeling of transport demand as soon as it comes to the practical 'translation' of these constraints via algorithms and models. The complexity arises as soon as individuals do interdependent activities and trips in a constrained environment, individual (or household) constraints being e.g. of physical or monetary nature. However, most often constraints in travel demand models are considered only in a reduced or simplified manner. For instance, a tour or activity-based model normally proves the feasibility of an estimated activity pattern regarding the time available for activity and travel, thus complying with the 'capability constraint'. But rarely will it accomplish with constraints of e.g. facility opening hours and even less with types of 'coupling constraints'. Nonetheless, beside the behavioral complexity, availability of data is another limiting factor. The more complex is the decision-making situation covered by a model, the more data is needed to provide reliable estimates on predictions of behavioral change.

Without doubt, many models and empirical studies indicate the strong correlation between principles of time geography and challenges that current models of travel demand analysis are facing. Generally, the relations between time geography and transportation research can be seen in the effect of time-space constraints on accessibility of choice options or on possible spatial paths (Miller, 2004, pp. 647-58). Timmermans et al. (2002) provide a classification of time/space-based approaches to travel demand. They differentiate between a) constraints-based models, b) utility-maximizing models, c) computational process models and d) micro simulations. In the nutshell, "constraints-based models have their roots in time geography, and utility-maximizing models in microeconomic theory, while computational process models have been inspired by psychological decision process theories" (ibid., p. 179). According to the same source, "these models (constraints-based) typically examine whether particular activity patterns can be realized within a specified time-space environment." Furthermore, "to examine the feasibility of a certain activity program, a combinatorial algorithm is typically used to generate all possible activity sequences" (ibid., p. 179). The latter is of relevance here, as the objective pursued in this thesis is similar aiming at 'bridging the gap' between the identification of time-space constraints and putting them into operation using a combinatorial algorithm.

A more recent approach to the concept of time-space constraints via so-called individual action spaces was proposed by Dijst and Vidakovic. They simulate 'potential action spaces' defined as "the area containing all activity places which are reachable, subject to a set of temporal and spatial constraints" (Dijst and Vidakovic, 1997, p. 120). According to them, the set of conditions includes activities types and locations as well as time resources. It is argued, given that there are two bases for activity participation, the general form of the potential action space is an ellipse (ibid., p. 121). Their work aims at developing a simulation to estimate activity locations to be reached within the action space. In addition, they raise the issue of combinatorics and computational limitations if all costs (e.g. times) between all feasible alternatives have to be calculated. Schönfelder and Axhausen (2003) extended the perspective of activity spaces by building up their estimation on data of a travel survey that lasted several weeks and by testing different activity space measurements. Cirillo et al. (2003) also made use of the concept of action spaces aiming at identifying the area where an individual is likely to conduct his activities. However, they focused on the generation of a choice set of alternatives pertaining to the action space to estimate behavioral parameters of a destination choice model. Newsome et al. (1998) support the elliptic form as a suitable approach to better understand travel behavior. They found that home location and household size are important factors in the determination of action spaces. For the purpose of analysis and visualization of time-space constraints different geographical measurements can be applied. In this context, the application of GIS became important to conduct "geocomputational algorithms for estimating spatial units that describe either the revealed or potential activity participation of individuals over space" (Buliung and Kanaroglou, 2006a, p. 36).

A preliminary conclusion from the review of recent work in the area is that the concept of time-space constraints stays crucial for the development of travel demand models. In many forms constraints are considered. A rather obvious example was that e.g. activity-based approaches try to assure that activities and travel times spent comply with natural temporal limits. A more elaborated example would be a model able to reproduce a joint activity arranged spontaneously between friends due to a relaxed constraint, say, because another activity took less time than expected. As was stated above, the identification of constraints and their inclusion into an empirical application is a major objective of the upcoming chapters. However, a look back at some significant statements of Hägerstrand's article is suggested before we close this chapter. Remarkably, their validity has prevailed for over 40 years of research in the area and their relevance for the objectives pursued in this thesis remains high:

- "Assume that each person needs a regular minimum number of hours a day for sleep and for attending to business at his home base. When he moves away from it, there exists a definite boundary line beyond which he cannot go if he has to return before a deadline. Thus in daily life everybody has to exist spatially on an island. Of course, the actual size of the island depends on the available means of transportation, but this does not alter the principle" (Hägerstrand, 1970, p. 13).

'Boundary line' or 'island' quoted here is equivalent to 'potential action space' defined by Dijst and Vidakovic or 'search space' in this work. The consideration of time-space constraints requires specification of these 'boundary lines'; respectively, the fulfillment of a constraints-based model implies the need to identify the (empirical) threshold values.

- "Every stay at some station means that the remaining prism (Note: i.e. the spatial area reachable given the time and means of transportation available) is shrinking in a certain proportion to the length of the stay" (ibid., p. 14).

This expression supposes that the time spent on a certain activity influences following decisions reducing room remaining for action. For any practical application it remains to verify and/or to quantify what 'shrinking in a certain proportion' actually means in values and how this could be considered at modeling of travel demand. In addition to Hägerstrands argument, it may be assumed that activity adjustments (temporally and spatially) occur in a way where the first activity of the day is adjusted because the second activity is less flexible (e.g. a work activity with a fixed time regime). This implies the revision of an individual's entire activity agenda along one day, including potential hierarchies between activities of different type.

- "Inability to rent a dwelling close to a place of work may, in the first instance, directly lead to long commuting times but may also lead to more concealed repercussions such as incursions on the time available for other activities" (ibid., p. 17).

This quotation lets us assume that the distance between home and work influences 'the time available for other activities'. To say in other words, there is a delimiting effect on additional activities with the increase of distance between home and work. This effect may embrace the time available for traveling to other activities and/or the duration of time spent there. Similarly to the previous citations, it remains to be defined whether these 'concealed repercussions' can be identified, quantified and eventually implemented in our model.

2.5 Summary

Pros and cons as well as main characteristics of trip-based, trip-chaining (tour-based) and activity-based model types were discussed. An additional section introduced main principals of time geography and related them to models of travel behavior. Similar methods are applied throughout the model categories such as regression models, (constrained) gravity models or multinomial logit models. Activity-based models represent an exception here, especially when individual's decision-making is simulated person by person via the use of behavioral rules. Data availability is also an issue and an important constraint for any of the approaches, but it was not discussed in detail since the focus was on the underlying methods. The common ground of all approaches is that demand generation leads to the provision of trip tables, i.e. trip matrices that can be assigned to networks of road and public transport. The discussion showed that still the main distinction is related to the modeling of individual or combined activities and trips. Another focal point was on discussing the paradigmatic change caused by the appearance of activity-based models and a more explicit consideration of time-space constraints in those models, first introduced by time geographers in the early 1970s.

There is no clear consensus about which approach might be indisputably 'the best' to estimate travel demand. The limitations of trip-based approach motivated the development of first tour-based and later activity-based models. The common conviction is that the theoretical concept of trip-based approaches is too simple and, consequently, is not sufficient for reproducing travel behavior appropriately. More ambiguous are opinions about the progress associated with the appearance of activity-based models and their implementation as simulations. Practitioners sometimes question their advantages. This does not concern the behavioral comprehensiveness of activity-based approaches, but their complexity and the missing financial support from the superior agencies, among other aspects, explain why the number of applications remains insignificant (Davidson et al., 2007, pp. 466-7). However, at least in the USA activity-based approaches are gaining territory and in some cases replaced traditional trip-based approaches (Davidson et al., 2007, p. 469; Bradley et al., 2010, p. 2).

The heading of the chapter 'Review of transport demand models' has been chosen intentionally, as the chapter focuses on the basic concepts underlying the models rather than on their detailed mathematical foundations. The chapter provides background for a better understanding of current approaches and gives insight into how the methodology to develop fits into the systematic of already existent models. Picking up some of the main features presented in previous sections, we summarize the following requirements for the model development described in the upcoming chapters:

- Travel demand is derived from the need to conduct activities, which means the basic unit of the analysis is an activity pattern. Respectively, our approach should facilitate the consideration of interrelated decisions in compliance with the characteristics of trip-chaining and activity-based approaches.
- The consideration of time-space constraints is crucial for the integration of more behavioral realism into the approach. Beside the identification of time-space constraints, the estimation of empirical values allowing for their quantification presents a particular challenge.
- The attempt is not to create a model that pursues the reproduction of individual decision-making but the behavior of aggregated groups (e.g. household clusters). However, the approach should be adjustable in such way, that theoretically any level of disaggregated group types can be considered (given the data available).
- The estimation of transport demand is intended to be represented by a probabilistic approach rather than by discrete choices of individuals. The estimation of fractional probabilities should be feasible for all considered decisions building part of an activity pattern and/or of a tour. In consequence a trade-off needs to be found between the level of detail required and the combinatorial complexity naturally coming up once decisions are modeled in an interrelated way.
- In consequence to the previous point, a crucial issue to cope with is the development of a (search) efficient approach to model destination choices. Obviously, aspects of mode-specific zonal attractiveness and accessibility should be considered. A solution needs to be found to define a reduced choice set that excludes remote destination options with very low probabilities. They rarely represent observed travel behavior decisions and increase significantly the number of choice options, thus increase the complexity to handle.
- The approach should allow – at least in the way it is designed – to accomplish equilibrium conditions, i.e. iterations with network models to update impedance values. This enables us to provide a single and stable solution to the complex problem of modeling (urban) transport demand.

We may admit that the named characteristics of the method we plan to implement have stayed rather descriptive so far. However, the objective was to outline the principal characteristics and requirements we think important before going into details of the analysis and application. To sum up, the former listing appoints to a model comprising both elements of trip- and trip-chain based approaches as well as the elements related to activity-based analysis. For instance, the aspect of household clusters and the estimation of probability flows show similarities to trip-based models; the consideration of interrelated decisions reflect the main characteristic of trip-chaining models, while the issue of temporal and spatial constraints clearly relates to aspects primarily treated in activity-based approaches. This actually describes the contribution of the approach coming up: it combines important aspects of different modeling directions staying with the paradigm of a single solution for the supply-demand interaction problem but allowing for more behavioral complexity than current trip-based and trip-chain based approaches.

If one had to provide a clear-cut definition of the approach developed in this thesis, it would be as follows: **'a probabilistic tour-based model considering time-space constraints'**. Interestingly, a clear statement regarding this type of model could be found in literature: "It proved to be impossible to apply a tour-based model on a regional scale in an aggregate fractional-probability fashion" (Davidson et al., 2007, p. 471). If we formulate this rather discouraging citation as a question, it would sound as follows: how should a model based on fractional probabilities look like provided it accomplishes requirements related to time-space constraints, to necessary behavioral detail and to manageable combinatorics?

The following chapters aim at further elaborating on the possibilities to handle this question methodologically and empirically. For this purpose, the formulated requirements are first embedded in a broader context of how travel behavior can be modeled introducing the concept of hierarchical choice-making. The following step will be the empirical application searching for the design of an efficient algorithm that would enable coping with the required behavioral detail and with the limitations imposed by the number of treatable combinations.

3 Methodology for the modeling of activity and travel pattern

The characteristics and requirements listed in the previous summary section, respectively their transformation into a model, call for a theory to interpret travel behavior and the related decision-making. Accordingly, basing on theoretical assumptions outlined in the following sections, here we focus on the introduction of the methodological framework. This description remains generic, i.e. no reference to specific models or data sets is made. Notice that the upcoming sections are partially resulting of common efforts of a working group aiming at establishing the theoretical foundations. Intermediate results were presented and discussed in several conferences related to this topic (see Martínez, Justen, Cortés, 2009a; Martínez, Justen, Cortés, 2009b; Justen, Martínez, Cortés, 2010).

3.1 The hierarchical structure of choices

The modeling of daily travel choices can appear extremely complex due to the large option set determined by the combination of activities including decisions for locations and for transport modes and the allocation of time. Such complexity imposes high requirements even for advanced computers (Iacono et al., 2008, p. 335; Davidson et al., 2007, p. 465). The number of options increases with the number of ways in which time is allocated, the set and order of activities are performed and with the level of geographical detail, i.e. the zoning system. At the same time, models are required to consider various types of constraints, at least in the sense of available resources, namely, time and money. In order to introduce a rational but also feasible strategy and to reduce complexity due to the combinatorial amount of options, the information needs to be organized into a hierarchical structure of decisions in time and space. The hierarchical structure is a plausible strategy as long as alternatives from different scales, i.e. hierarchical levels, are substantially different in their expected utility or regarding the consumption of resources that constrain the consumer's rational choice domain.

To illustrate how hierarchies are supposed to occur in transport decision-making a following example is used. We assume that a new shopping mall opened next to a residence block. This probably alters the shopping routines of individuals, as some stores that used to be frequented lose in attraction since the new mall offers the same or even better options. However, it appears unlikely that the new mall will affect the work regime, i.e. the time leaving for work and the time staying at work. The conclusion is that the new shopping mall will have an impact on the decisions of where to shop and, perhaps, on when to shop provided the mall has longer opening hours. This simple example illustrates that change in the supply structure (here the new shopping mall) does not necessarily affect all daily travel choices but only those that are related to the hierarchical level of shopping-related decisions. In consequence, it might be postulated that some behavioral regimes remain stable (Buliung and Kanaroglou, 2007, p. 153-4; Pendyala et al., 2002, p. 74). On the other hand, if work-related decisions change (due to a new job location or adjusted starting time and duration), other decisions of, e.g. where and when to shop afterwards, will most likely be affected and adjusted. One can conclude that superior decisions (here related to work activity) influence the subordinated choices (related to shopping activity) while those related to the subordinated level do not affect the superior decision level.

The hierarchy described in this example reflects grades of importance attributed to different activities. While some activities, like leisure or shopping, show a high degree of flexibility, others, like work or education are more fixed and form an essential – assuring individual subsistence – part of life. It may be argued that leisure or shopping can also be compulsive activities, especially if work or education are not the part of the individuals' activity and travel agenda (Kitamura, 1997, p. 15). However, the issue of priorities and hierarchies between activities comes up even if no clear hierarchy is visible at first sight. If three or more leisure activities are conducted on the same day, it appears reasonable to assume that one of them has given the initial impetus for the person to leave the house. If so, how did the decision for the principal leisure activity influenced the adjacent ones, regarding location and mode choice or starting time and duration of activities? Unfortunately, the information available in transport planning and modeling often does not allow to answer these questions unambiguously, since most travel surveys only reveal what actually happened without exploring the motivations for or interdependencies between decisions (this discussion is continued in more detail in section 4.3.2).

The existence of hierarchies in transport-related decision-making appears to be valid both in the temporal and in the spatial contexts. For instance, regarding the spatial dimension it is unlikely that an individual searches over the entire space for a location for, say, resolving a private errand. The cognitive capacity of an individual is limited and not able to evaluate all theoretically feasible alternatives; naturally, such an effort would be highly inefficient and unlikely regarding the (constraint) time resources available. One can expect that a person tends to apply a hierarchic, i.e. selective search where only a processible number of alternatives and their characteristics are taken into account. In fact, there is broad and long-lasting evidence available regarding the limited processing capacity of the human mind where authors suggest a memory span of about four to ten alternatives (Malhotra, 1982; Miller, 1956). In this context the concept of hierarchical decision-making appears to be twofold: first, the set of potential locations for e.g. the private errand activity will be a subset of the total number of options available. Second, locations are likely to depend on prior location decisions taken the same day, especially those related to repeated activities because the spatial knowledge about this area is likely to be better than about areas traveled to less frequently.

Concerning the temporal dimension, hierarchies can be observed at different levels (see Table 3.1). For instance, decisions related to infrastructure, construction or technology development are taken for a very long term perspective and condition – at least partially – decisions of e.g. where to look for a residence location or whether to purchase a car or not. Thinking one level further, the level-of-service provided by infrastructure and public transport operators determines decisions taken in the shorter term for regularly visited activity locations and transport modes. Finally, daily circumstances of the operational level or changing weather conditions affect the decisions in the very short term, e.g. decisions made about local routes, modes or locations. In addition, temporal hierarchies occur also at a single scale, for instance, regarding the time spent on different activities. The time spent at work conditions the time available for any other activity conducted on the same day. If more than one tour is made one can expect that time dedicated to the most important tour (containing the work activity) influences the time (and the related decisions) available for any other tour and so on (Bowman and Ben-Akiva, 2000, p. 7). What is common for all the decisions independently of the hierarchical level is that they have to happen within boundaries defined by monetary or time-depleted resources. Table 3.1 suggests a time hierarchy for activities considering the spatial dimension for location choices in the long- to short-term perspectives. In the table a further differentiation is made between decisions taken by consumers (i.e. population) and those taken by suppliers (i.e. decision-makers or operators).

Table 3.1: Time hierarchies for activities

	TIME DIMENSION	TIME UNIT	CONSUMERS	SUPPLIERS
1	**Very long term**	Centuries	Cultural factors	City structure: re-building and land use pattern New cities
2	**Long term**	Decades to Years	Family structure Education Car ownership Residence location Work location	Infrastructure: Buildings Roads/Tracks Bridges Technology
3	**Short term**	Years to Months	Time and location of discretionary activities Transport Modes	Operations: Level-of-Service (LOS)
4	**Very short term**	Days to Hours	Route Choice Local Transport Modes Walking Destinations	Daily LOS, cicumstantial factors (e.g. weather)

One principal conjecture is that time dimensions are upward dependent, i.e. decisions related to the short-term period depend on decisions related to the long-term period and so on. Notice that hierarchical dependencies may occur in one time dimension, too. In some cases this is more evident, e.g. the duration of a leisure activity depends on the time spent at work, in other cases mutual dependency may occur. For instance, whether the choice for a residence location depends on car ownership or vice versa is less clear. However, it becomes apparent that transport system cannot be treated isolated from the entire urban system, specified in the levels one and two. The appearance and structure of any city and its transport system can be attributed to a very long-term influence of a variety of cultural factors, for instance including economy, religion and political beliefs. Any subsequent decision-making can be related to a subsequent time dimension, respectively. Similarly, the long-term activities specified in level two exert influence on the activities in lower levels.

In modeling practice it can be observed that long-term decisions (see level two) are often treated and modeled separately from the decisions belonging to daily activity and travel behavior (see level three) (Iacono et al., 2008, p. 327; Davidson et al., 2007, p. 473). Thus, parameters like car ownership in households or work and home locations, once assigned or modeled, remain stable at least for a time period (within a year or a day), which a transport demand model is applied for. For the analysis of level three and four (short to very short term) travel demand models are applied as described in the previous chapter. Even in the lower levels of short- to very short-term decisions, the issue of time and spatial hierarchies persists. The inherent degree of detail on level four might often exceed the capabilities of the travel demand models discussed before. Particularly, the required spatial fineness of zones considered may only be provided by microscopic traffic simulators, which for the sake of detail cover only a spatial fraction of the area under investigation. However, the route choice related to the very short-term decisions once again reflects our argument of a limited cognitive capacity to process choice options by the individual. Similar to the prior discussion about its effect on destination choice it is likely that an individual actually knows only a reduced set of routes rather than all practically feasible ones.

To sum up, the Table 3.1 accomplishes two functions: first, it puts the concept of hierarchies in transport decision-making into a broader context of time and space. This allows relating the levels and processes to different (transport) models available. Second, it allows referring the method developed and applied in this thesis to a respective level. The approach presented here is related to level three and represents an attempt to model daily activity and travel behavior in a constrained environment. With regard to the hierarchical concept, decisions of level two are taken into account and are integrated as a given input but are not modeled explicitly. Within the third level further hierarchical dependencies between locations, modes and times are considered. These issues will be discussed in the following section. Regarding the relationship to the lower level four, the demarcation is less clear. Generally, we might say that the generic approach as presented in this thesis is applicable on a more detailed and also more long-term level as long as the respective information is available.

3.1.1 The concept of macro and micro processes

The concept of hierarchical decision-making is now further discussed from the two principal aspects of time and space. We assume a differentiation between two hierarchical levels, an upper, macro level and a lower, micro level, while both terms of macro and micro apply for temporal and spatial aspects. There exist not only hierarchies but also dependencies between the temporal and spatial scales as well as among the scales themselves.

Hierarchy among temporal scales: Long-term decisions (macro scale) influence decisions taken in shorter term (micro scale). Another example related to Table 3.1 can be the decision to purchase a car, which would in turn influence subsequent mode choices. The decision to buy a car is taken at a macro scale (sporadic, in the long term), whereas once the car is available it influences the mode choice (regularly, in short to very short terms). Beside hierarchies between different levels, the concept of macro and micro processes can also be applied within one single time dimension (e.g. a day). For example, time spent at work ascertains the duration of following activities. Hence, the work duration belongs to the category of a macro process because it is a result of a long-term time investment (visiting school, university, etc.), while the duration (and the relative time investment) related to discretionary activities is subordinated, thus representing a micro process. Along with it, a macro process has certain inertia in adjusting to new circumstances, while a micro process can adjust faster. Naturally, in both cases (macro and micro levels) decisions might be conditioned or constrained by limited time and/or money.

The latter point is of vital importance for the approach developed here as the scale of each process is defined by the amount of resources required. This principal conjecture can be exemplified like this: bigger time and money resources are needed for a macro decision (e.g. purchase of a car or change of residence or work location) than for a micro decision (taking a car instead of a bus). Our 'resource argument' becomes cogent thinking in the time normally needed to prepare and/or plan the respective activities, especially those of the macro scale. The strength of the argument to define scales in accordance to the resources invested lies in its simplicity and traceability. So far we present this as one of the main theoretical conjectures but later on introduce the respective empirical evidence.

Hierarchy among the geographical scales: Long-term macro decisions on residence or work locations determine the subordinated micro decisions of a short- to very short-term character. Activities inert in their adjustment 'define' the spatial structure in which subsequent decisions for modes and locations are realized. One can illustrate this with the following assumption: the spatial axis between home and work locations (macro level, long-term period) and the knowledge about the potential options between them determine the subordinated location choice (micro level, short-term period). Moreover, one single location can be related to two levels of spatial hierarchies, i.e. to two aggregation levels. In Santiago, the macro level is spatially characterized by macro zones and comprises 618 Traffic Analysis Zones (TAZ), while the micro level is a further disaggregated zone system that comprises approximately 50.000 residential blocks (see Figure 4.1). Given that each micro zone is located completely within one macro zone, each micro location belongs to both the micro and the macro levels.

This set of rules dictates the general hierarchical structure of individuals' choices in urban systems, which has the following implications for building of a hierarchical model: the temporal scale defines long-term effects of macro scale choices both in terms of the consumption of resources and of the level of utility attained. The appropriate geographical scale is subordinated to these decisions and is considered adequate for implementing a more efficient complex search process, which would save (computational) efforts and resources. This implies that this search is done not through all theoretical options but through a subset, which remains to be defined. At a macro level the spatial search, e.g. search for work or for a job location, may cover the entire city, while the subordinated decisions on shopping or leisure include a spatially reduced set of alternatives. Hence, macro-level processes consume and produce a larger amount of individual resources than micro-level ones. The implications for the methodology are twofold: first, we can assume that the application of the hierarchical concept of travel behavior is more realistic; second, it is expected that the search process can be designed in an efficient manner, reducing the number of combinatorial options and saving computational resources.

3.1.2 Hierarchies of decision-making in daily travel

The main characteristic of the methodology developed is that it is based on a probabilistic formulation of transport demand. Hence, the objective is to be able to estimate so-called spatial path flows, which reflect the joint probability of a set of decisions made on the choice of an activity pattern or of a tour, on locations, on modes and on times. Below we introduce the choices considered in the methodology, provide a comprehensive terminology and present a detailed step-by-step description of the generic framework.

The previous sections allow concluding that the spatial path (i.e. sequence of activities and trips) can be decomposed into a hierarchical structure of conditional sub-choices, constrained by the total of resources of an individual. We focus on a bi-level of hierarchies whereas the bi-level applies for activities, locations, modes and times. The approach is probabilistic, where the highest probability reflects the highest utility attained for the most probable sequence of choices. Additionally, an individual is assumed to save resources at a macro level to be able to make a different choice on a lower micro level; thus, decisions are conditioned by the remaining resources. In practice this means that a change in probabilities related to a subordinated decision, e.g. location of a leisure facility depends on an alteration at the superior level, for instance, reduced work duration. This hierarchical approach assumes that individuals first decide on the long-term activities of an activity pattern, and only after that they make decisions on activities and trip-related choices that are referred to the next (micro scale) level.

The conditionality between temporal events in transport models is sometimes treated making use of so-called hazard-based models. Essentially, with these models one tries to predict the probability that an activity ends given the time it already lasted (Hensher and Mannering, 1994, p. 64; Bhat and Pinjari, 2008, p. 105). However, these models require comprehensive information (often stated-preference data) about the individuals actual temporal decisions as well as about relevant choice alternatives to be able to estimate model parameters in a discrete choice framework (Ettema et al., 2007, p. 835). To face the assumption that this kind of data is rarely available, we define conditionality between durations directly from the observed activity patterns. This means that e.g. the macro activity duration depends on the macro starting time, while the micro starting time depends on the macro duration time.

The previously said allows to assume the following: a) decisions on any activity at the macro scale are made right away at the beginning of the macro-scale period and remain unchanged during this period; b) at any time, the choice of a feasible path (sequence of choices) is conditioned by budgets of time and by the activity pattern chosen at the macro scale. Consequently, behavior related to short-term, micro-scale choices, is conditioned by the longer-term choices. Another consequence is that the same individual or a user group may choose different spatial paths depending on the set of choices this individual or user group has already made on a macro scale. In general, this conditional choice process also includes durable goods, such as vehicles or apartments, which affect short-term decisions taken in the transport system like choices of travel modes, of locations and routes of short-term activities.

However, as we focus on short-term daily travel behavior the long-term decisions on work location or on purchase of a car are taken as given and do not form part of the methodology presented. This does not exclude their influence on travel choices in the sense that they may alter and affect the subordinated choices.

The solution to the problem of conditional choices is a joint probability where each choice probability is dependent on the choice probabilities of activities at larger time scales. The same applies to the geographical scales. Short-term location decisions are conditioned by activities and locations decided upon at longer-term scale. Potential locations are those that are feasible, considering the activity and travel time already spent. The search of feasible locations is constrained by the time and cognitive resources of an individual. In any case, a joint choice of activities and of trips – a spatial path – must be feasible against a set of criteria applied. The independency of the macro scale from the micro scale and, conversely, the dependence of the micro scale on the macro scale suggest a multiplicative property of the methodology calculating conditional probabilities. The following Figure 3.1 summarizes the concept of macro- and micro-scale choices, putting it into the context of travel demand decision-making.

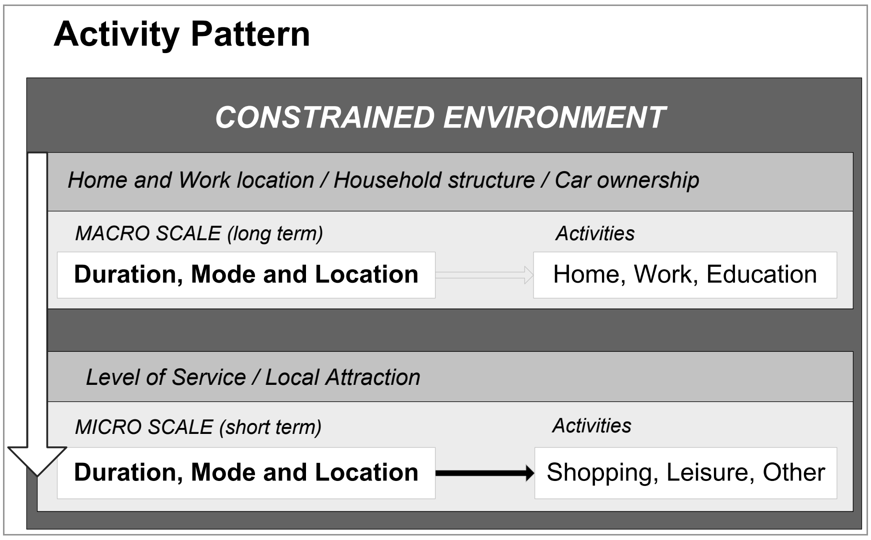


Figure 3.1: Macro and micro scale: sequence of choice processes

Annotation: An activity pattern describes the number and sequence of activities and trips conducted during a given time period, see 3.1.3)

Figure 3.1 specifies the idea of hierarchies in time and space as introduced before and relates it to the level of daily travel behavior decision-making. Respectively, the starting point for the estimation of a spatial path is the choice of an activity pattern. Within each activity pattern there is a differentiation between macro- and micro-scale activities. Activity duration, mode and location are first assigned to the out-of-home macro-scale activities like work or education. It is important to realize the difference of influence factors at each scale: the long-term impacts are to be found at the upper level, and the short-term impacts occur at a lower level. For instance, a new job changes decisions related to the upper, macro scale, while a new shopping mall changes those related to the lower, micro scale. The arrow from the macro down to the micro scale – and not vice versa – indicates that changes at the macro scale, both temporal and geographical, constrain the micro scale decisions. Although this concept is applied against the background of daily travel behavior, theoretically, the bi-level model can be expanded to further levels as long as it appears reasonable to treat behavioral choice domains independently. It is important to underline one more time that all the processes occur at any level in a constrained environment. Generally, these constraints are related to time, space or finances, while some constraints can also reflect social interactions within households or between them.

Within this section the general distinction between macro and micro processes was set into the context of transport demand analysis. It is important to recognize that this approach differs from sequential models that estimate activity and travel flows in a consecutive order. In difference we suggest a hierarchical approach where activities (and related decisions) are assigned to macro and micro scales. The micro scale activities are adjusted according to the macro scale decisions, thus we treat macro scale choices independently. This macro-micro differentiation is maintained throughout the development of the generic framework as well as all the way through the empirical application later on. These remarks have remained on a conceptual level so far. The next section will introduce a detailed terminology followed by a concrete step-by-step methodology aiming at modeling activity and travel patterns.

3.1.3 The hierarchical tour – terminology

The introduction and use of a clear terminology is indispensable to avoid ambiguity in description of terms and processes. In the field of transport modeling some terms are used as synonyms to describe one process. For instance, a sequence of activities and trips, which is a general outcome of travel survey, may be denominated as activity pattern, activity set or activity chain. With the objective to avoid ambiguousness we provide an overview of terms that are used in the subsequent step-by-step description of the methodology as well as in the explanation of the empirical application given in later chapters.

Macro and micro scale: the terms are used both in the geographical and in the temporal contexts; a *macro scale* can be disaggregated into an integer number of *micro-scale* units, i.e. a *macro* period or zone contains *micro* periods or zones; a *micro scale* never overlaps a *macro scale.*

Trip: the term defines movement between origin (O) and destination (D) at a given geographical *macro* or *micro scale.*

Location, Destination: both terms are used synonymously and describe the geographical area where an *activity* takes place; *locations* and *destinations* on a *macro scale* are represented by Traffic Analysis Zones, locations on a *micro scale* are examined at the level of residential blocks.

Activity or trip purpose: both terms are used synonymously; *activity* and *trip purpose* describe the action realized at a *location* or *destination* of a trip.

Tour: the term stands for the combination of *activities* and *trips* that start and end at the home location; a *tour* contains one or more out-of-home *activities*; it does not include information about modes used or about temporal aspects (i.e. starting time and duration of *activities*).

User groups (UG): the term specifies the pooling of the population into socio-economic household groups; each *user group* is defined by the number of motorized vehicles (0, 1 and 2 or more) and income available per household (see Table 4.1).

Activity pattern (AP): the term comprises one or more *tours* and describes the number and sequence of *activities* and *trips* conducted during a given time period (a day); analogous to *tours*, information about modes used and temporal aspects is not considered; if an *activity pattern* consists of only one *tour*, the terms *activity pattern* and *tour* can be used as synonyms.

Primary activity (PA): the term describes an *activity* of the highest priority according to resource consumption and the expected utility attained; according to the hierarchical concept, the *primary activity* belongs to the *macro scale*; in a *tour* or *activity pattern* related to work or education these are considered *primary activities*; in a *tour* or *activity pattern* related to other purposes the *activity* of longest duration is considered *primary activity*; the home location belongs to the *macro scale* and is referred to as a *primary activity* as well; 'X' denotes a primary activity.

Secondary activity (SA): the terms describes an *activity* of subordinated priority according to resource consumption and the expected utility attained; according to the hierarchical concept, the *secondary activity* belongs to the *micro scale*; given the definition of the *primary activity* at the *macro scale*, any other *activity* is of secondary priority; 'Y' denotes a secondary activity.

Activity pattern type (APT): the term means the aggregation of similar *activity patterns* into *activity pattern types*; the *primary activities* of work and education are denominated with letter 'X', the *secondary activities* with letter 'Y'; for instance, the *activity pattern* of Home-Work-Leisure-Home and Home-Education-Shopping-Home are grouped into Home-Primary-Secondary-Home or 'HXYH', respectively.

Activity pattern probability (APP): the term defines the probability that *activity pattern* are conducted by the household *user groups*.

Pattern demand (PD): the term defines the product of the number of households by *user group* in a *macro zone* and the *activity pattern probability.*

Detour factor (DF): the term denominates the geographical detour made due to a *secondary activity*; it is calculated as the ratio of the distances between locations of *primary activities* and additional distances to the location of the *secondary activity*; DF is used for the estimation of the *search space* (see next term).

Search space (SS): the term characterizes the geographical area and, thus, the number of *macro zones* surrounding two *primary activities*, i.e. the locations of home and work or education; the *search space* includes all feasible zones where a *secondary activity* may be conducted; its geometrical manifestation is an ellipse based on the *detour factor* as specified above.

Land-use density (LUD): the term defines square meters of built commerce and service supply facilities per square kilometer of *macro zone* area.

Land-use attractiveness (LUA): the term defines the ratio of *land-use density* by zone to the summed *land-use density* by *search space.*

Local attraction factor (LAF): the term describes the combined attractiveness of each *macro zone* belonging to the *search space*; it is defined by the *land-use density* divided by the travel time by mode needed to reach the *macro zone.*

Location weighting parameter (LWP): the term describes parameter for the adjustment of the *local attraction factor* in order to reproduce empirical travel time distributions towards *secondary activities* by mode.

Final choice set (FCS): the term describes the selected subset of *macro zones* within the *search space* according to the product of *local attraction factors* and *location weighting parameters.*

Mode Combination (MC): the term defines a sequence of modes used throughout a *tour*, where each *mode combination* is characterized by the main mode, i.e. the mode used to travel between the locations of *primary activities*, and by further modes used to connect the remaining *secondary activities.*

Daily travel time (DTT): the term describes the sum of travel time of all *trips* conducted within a *tour* or an *activity pattern.*

Spatial path flow, (SPF): the term comprises the *activity pattern* or *tour,* the *macro zones* visited, the modes used (thus, the *mode combinations*) as well as the starting time and duration of all *activities* and *trips*; the *spatial path flow* is obtained by multiplying *pattern demand* by probabilities of performing *primary* and *secondary activities.*

3.2 The hierarchical tour – methodology

In this section reference is made to the theoretical deliberations introduced in the previous sections and illustrated in Figure 3.1. The activity pattern type, which serves as an example to illustrate the translation of the theory into a practical application, is a tour of type Home-Primary-Secondary-Home (HXYH). The selection of a tour that combines two out-of-home activities (with one being of primary importance) is meaningful to show how macro and micro scales are integrated into a single approach. The methodology aims at providing an efficient solution that would enable modeling an activity pattern (two activities, their starting time, duration, and location choices) and a travel pattern (three trips and the related mode choices).

The method departs from the calculation of fixed probabilities provided by a trip-based (macro scale) model for the Home-Work relation and the associated mode decisions. For the subsequent decisions related to activity 'Y' (micro scale), probabilities are calculated directly from data. The objective is to present a generic method, which means that no reference is made to the actual data sources or models used later during the application example; this description serves rather as a stand-alone result offering a potential for application in any context.

STEP 1: ACTIVITY PATTERN PROBABILITY AND MODE COMBINATIONS

First of all, some input information has to be provided. We assume that both the Home-Work relation and the associated mode choices for reaching the work location are known. This information is assumed to be further disaggregated by socio-economic characteristics, for instance, by household user groups. Generally, a trip-based transport model can provide this information. It is denominated as:

- Probability to travel from Home to Work by mode and user group.

As a following step, we have to specify travel demand in terms of number and type of activity patterns conducted. Note that the activity pattern taken as an example is of type Home-Work-Secondary-Home, with the work activity representing the 'X' in 'HXYH'. In addition, information about how many households by user group conduct the activity pattern during one day is required. Generally, a travel survey provides information about activity patterns conducted by socio-economic groups; thus, the following combined probability is defined as:

- Joint probability of the number of households by user group and the probability that they carry out the activity pattern of type Home-Work-Secondary-Home; i.e., the pattern demand (PD).

The result of this probability is similar to the initial demand-generation step realized in aggregated travel demand models. There, the average number of trips by time period and behavioral homogeneous group is multiplied by the total of population by group. For instance, the respective share or probability says that a person aged between 30 to 40 years, employed and having a car conducts on average 0.6 trips per day between Home and Work locations. In the approach described here this procedure is expanded to the activity pattern perspective. In result, the share (or probability) is estimated that, for instance, a household of one of the user groups follows the respective activity pattern on average 0.03 times per day.

As it was mentioned above, the mode and location choice probabilities for the work location are assumed to be set by a trip-based macro model. Depending on these main mode probabilities that define the Home-Work relation, the mode combination probabilities – i.e., the modes chosen for the remaining trips to the secondary location and back home – are obtained from empirical observations and are defined as:

- Mode combination (MC) probability by user group.

It is assumed that these probabilities can be generated from a regular travel survey. In the approach described here, they are determined once and remain static during the application (see the respective empirical results in section 5.1).

In result it is now known how many households by user group realize the activity pattern, to which location and by what mode they travel to work and what mode combination probabilities are given according to the mode choice for the primary activity. So far, the destination choice for the secondary activity is not defined yet.

STEP 2: STARTING TIMES AND DURATION OF ACTIVITIES

This step defines the expected starting time and duration of all the 'X' and 'Y' activities. It can be assumed that temporal aspects depend on the individual's situation, i.e., on his or her employment status (full time or part time) or on the conditions set by the household structure. To capture potential influences of this type, it is assumed that temporal aspects differ by socio-economic user groups. Respectively, the following information needs to be provided:

- Starting time (S_W) and duration of the primary activity (D_W), both indicated on an hourly basis;
- Duration of the secondary activity (D_Y) indicated on an hourly basis.

According to the assumption of macro and micro scale choice dependencies, the starting time and duration of the secondary activity depend on the decisions taken for the primary activity. The starting time of the primary activity depends on its duration. This dependency ensures that, for instance, a long-lasting primary activity of, e.g., 13 hours is unlikely to start very late in the day (see section 5.3 for empirical evidence). In other words: the probability that the primary activity starts late in the day declines with its increased duration. In addition, a long duration of 'X' leaves less time available for 'Y'. The associated probabilities cannot be considered in a fully disaggregated manner. This means that not every spatial path contains all theoretically feasible combinations of starting times and duration. In the example treated here and based on an hourly aggregation of time, this would lead to 24^3 (13,824) options of starting and duration time combinations (24 feasible hours to start activity 'X', 24 possibilities that activity 'X' lasts 1 to 24 hours and 24 feasible durations for the 'Y' activity). First, the number of theoretically feasible temporal combinations can be substantially reduced to a set of combinations observable in reality. Doing so, we exclude highly unlikely events as, for instance a 24-hour lasting activity 'X', followed by a 24-hour lasting activity 'Y'. Second, of the remaining set of conditional probabilities only one single temporal combination is assigned to each person, i.e. to each socio-economic user group at the origin.

The assignment process is organized as follows: Notice that each origin has to be disaggregated by the number and by type of socio-economic user groups considered. Then, the distribution of the primary activity duration in hours and by socio-economic user groups is determined. Respectively, the frequency distributions represent a probability that, for instance, the 10-hour duration of activity 'X' occurs with 20% probability, the 5-hour duration with 7% probability and so on. Subsequently, a random number between 1 and 100 is created and assigned to the origins disaggregated by user groups. In relation to random numbers, the duration of activity ‘X’ is now assigned considering the frequency distributions. This assures that on average across all spatial paths, frequency distributions observed in the empirical study are reproduced with the model. The same method is applied to define the starting times, where probabilities of starting at a certain hour depend on the duration of activity ‘X’. Again, a random number between 1 and 100 is created and the starting time is assigned in accordance with the previously established duration. Eventually, probabilities of the duration of secondary activity are assigned depending on the duration of the primary activity. One more time, a random number is created and the duration of activity ‘Y’ is assigned in accordance with the frequency distributions of its duration. This sequential process, which considers the hierarchical dependencies between choices, ensures that timing aspects of the activity pattern are included.

At the end of this step, the starting times and durations of activities ‘X’ and ‘Y’ are defined.

STEP 3A: DEFINING SEARCH SPACES – THE SPATIAL CONSTRAINT

The definition of search spaces is central to the approach. Their application follows the assumption that given an OD relation of Home-Work not all potential options to conduct activity ‘Y’ are taken into account. It was argued that this is due to the limited cognitive capacity of an individual to evaluate all the options and the limited resources available. The temporal and spatial constraints are used for defining the number of options that appear feasible. First, a spatial constraint is applied.

- Spatial constraint: detour factor.

Detour is defined as a maximum spatial deviation from the axis between home and work locations to reach a secondary activity location. This constraint implies that an individual (or a household) searches within a constrained geographical environment, i.e., between a confined number of zones. The detour factor is dependent on the time spent on the primary activity, the distance between home and work locations and allows for the geographic reduction of the choice set for the secondary location (the empirical analysis of detour factors is subject of section 5.4). The latter implies that for each OD relation of Home-Work several detour factors have to be estimated given the differences of time spent on activity ‘X’. The time spent on the primary activity depends on socio-economic characteristics of a person or of a household. Thus, the detour factor is indirectly also dependent on socio-economic attributes. Notice that, theoretically, other attributes – for instance, a different urban environment or a different set of socio-economic characteristics – may be taken into account to define detour factors. The actual computation of the search spaces can be done by applying the detour factors for the estimation of ellipses around home and work locations (see section 6.2.1 for the explanation of the ellipse generation process). In practice this means that for each Home-Work relation, several ellipses are determined leading to a set of search spaces, thus, to distinct (but reduced) numbers of potential secondary locations.

STEP 3B: DEFINING SEARCH SPACES – THE TEMPORAL CONSTRAINT

Due to the effect of the spatial constraint, a reduced number of location options for a secondary activity remain. Now, for each combination of zones (Home-to-Work-to-Secondary Activity and back Home) the respective impedances (times; if applicable, costs) are attached. In summary these are:

- Travel times (if applicable, costs) by mode for the first trip leg of OD relation Home-Work;
- Travel times (if applicable, costs) by mode for the second trip leg of OD relation Work-Secondary Activity;
- Travel times (if applicable, costs) by mode for the third trip leg of OD relation Secondary Activity-Home.

Subsequently, the temporal constraint is applied and defined as:

- Time constraint: maximum daily travel time by mode combination.

In practice this means that a total daily travel time is calculated only for those zone combinations that are equal or below empirically defined thresholds (for the empirical results refer to section 5.2). Respectively, the temporal constraint further reduces the set of zones attributed to the search space depending on whether these zones are reachable given the empirical thresholds for the maximum daily travel time by mode combination.

With this step a set of potential locations for secondary activities is defined. For each OD relation, household user group, mode combination and distance category of home and work, a separate set of secondary activity locations is identified in accordance to the time-space constraints.

STEP 4: ATTRACTIVENESS – THE LAND-USE

The locations assigned to the search space are not yet weighted regarding aspects of accessibility or attractiveness; thus, they are 'equally important'. Once the impedances in travel times are assigned for each potential secondary location the attractiveness using a land-use indicator is defined. The description of the land-use depends on the specific secondary activity modeled, for instance, for a shopping activity one requires information about commerce attractiveness of the location, for a leisure activity information about the amount and size of respective facilities or green areas. To keep it general at this point, the land-use attractiveness is defined as

- Square meters of land-use by square kilometers of the destination area size, i.e., land-use density (LUD).

Land-use is used as a measure of density. Thereby, the effect of differently sized geographical units – e.g., smaller zones in the city center typically get larger towards the periphery – is eliminated (see Figure 4.1).

With the inclusion of a land-use variable, a measure of attractiveness is introduced for each secondary activity location.

STEP 5: RANKING LOCATIONS WITHIN THE SEARCH SPACE

In the following, criteria are introduced aiming at ranking choice options within the search space. This generally applies to the land-use of each secondary location (attractiveness, step 4) and the travel time by mode required reaching the locations (accessibility, step 3b). At a glance, the criteria are defined as follows:

- Land-use density by zone in proportion to the summed land-use density by search space, denominated as the land-use attractiveness (LUA);
- Travel time to the secondary activity by mode.

Firstly, the land-use attractiveness is calculated by dividing the proportion of each zone by the summed land-use of the respective search space. Secondly, each proportional share is divided by the travel time required to reach the secondary activity location. This factor is defined as a local attraction factor (LAF, see 3.1.3). Naturally, zones with high land-use attractiveness and within short distance either to home or to work locations are ranked in front positions. This becomes a problem if a final choice set is ascertained based on the ranking. If only a limited number of alternatives is chosen, it is likely that options within short distance are overrepresented. To control this phenomenon, ranking via the local attraction factor is done according to travel time categories. Thus it is taken care of that a predefined number of choice options enters each travel time category (e.g. 3 options in at maximum 10 minutes travel time distance, 2 options in 11 to 20 minutes travel time distance and so on). This allows incorporating highest ranked zones by travel time categories, thus, considering location options, which are situated further from primary activities of home and work.

With step 5 all feasible secondary activities locations are ranked according to their attractiveness and accessibility.

STEP 6: DEFINING THE FINAL CHOICE SET

The application of time-space constraints already implied a substantial reduction of the initial choice set, which without any constraint would have been the entire city. However, many feasible options may remain. The definite number of options, i.e., the final choice set (FCS), is assumed to be a subset of the options resulting after application of constraints. A further reduction of options is meaningful, as it does not appear rational to assume that people search and evaluate options over the entire (search) space. According to Miller (1956), the maximum amount of options to be taken into account are seven +/- two (see also Malhotra, 1982 for evidence in marketing and Caussade et al., 2005 for evidence in transport). We adapt to these findings and consider at maximum seven choice options per primary location. This is important as final choice sets are defined both for the home and work location. The reason to that is that we assume that a tour can be either home or work related, i.e., that the location of ‘Y’ is either in proximity to the home or work location. The tour relation is defined as follows:

if the spatial distance between activities ‘X’ and ‘Y’ is longer than between ‘Y’ and ‘H’, the secondary activity location is considered as home-related, otherwise – as work-related (further explanations and empirical evidence to this assumption is given in 5.5.2). The selection process of alternatives is then based on the ranking done in step 5. Generally, highest ranked options according to the local attraction factor and differentiated by travel time categories enter the final choice set.

In addition and still before the definition of the final choice set we introduce a random process. The motivation is to take care off that not in any case the objectively ‘best’ (highest ranked) option is the actual choice. We assume that choices are taken within a range of uncertainty, as it is unlikely that the individual (or user group) is fully aware of the ‘objective’ attractiveness of all alternatives. Neither is it rational to assume that we capture via the local attraction factor the entire quality, i.e., attractiveness of a location. The random process is employed in such way that values estimated for LUA are multiplied several times with a random factor adopting values between 0.5 and 1.5. Any decimal value between 0.5 and 1.5 for the random factor is possible as well. We choose this range to control that the LUA are at maximum halved or gaining plus 50% of their original value. In consequence, we calculate several LAFs, obtaining different rankings of alternatives. We then choose those TAZ for entering the final choice set that most often represent the ‘best’ (highest ranked) options. Remember that for both the home and work location seven alternatives are chosen and that we consider whether a tour is home- or work-related and the observed distribution of trips by travel time category, such that the percentage of options in the choice set replicates observed percentages of trips by travel time category, for instance provided by a travel survey.

STEP 7: CALIBRATION OF THE DESTINATION CHOICE

After defining the final choice set, an adjustable weighting parameter is introduced. Denominated as the location weighting parameter (LWP), it serves to calibrate the secondary activity location choice aiming at the reproduction of empirically observed travel time distributions. In practice, this adjustable factor is multiplied by the local attraction factor, thus, altering the probability that a secondary activity is chosen. In consequence, this affects the spatial path flows (see step 8 below). The final parameters are the result of an iterative calculation process, where the criterion for halting the calculations is a minimum deviation between estimated and observed distributions. At the beginning of the calculation the parameters are set to a default value of 1 and are then sequentially adjusted.

The result of step 6 is a final choice set for secondary locations of no more than seven alternatives per primary location. We defined the final choice set including the application of a random process covering the aspect of imperfect knowledge. Finally, in step 7, we search for appropriate adjustment parameters to match estimated with observed trip distributions of secondary location choices.

STEP 8: CALCULATION OF SPATIAL PATH FLOWS

In this final step the estimated probabilities of previous steps are multiplied together. Accordingly, these are:

- Joint probability of the number of households by user group and the probability that the 'HXYH' activity pattern will be carried out;
- Probability to travel from Home to Work by mode and user group;
- Mode combination probability by user group;
- Probabilities related to the zones representing the final choice set.

Step 7 and 8 are applied in an iterative manner. Notice that probabilities of the set of zones that represent the final choice set have to be normalized to 1 (or 100%) after every iteration. In the case of a mismatch between estimated and observed trip distributions, say for instance too many trips are estimated to occur in the time slot of 1 to 10 minutes, in a second iteration zones within this category receive a location weighting parameter below 1 to reduce the spatial path flows. This process is applied for all travel time categories by mode and is repeated until empirical and estimated distributions approximate. Hence, the location weighting parameters are adjusted (step 7) and the spatial path flows are calculated once again (step 8).

The final results are spatial path flows. They refer to the probability that an individual of any user group conducts a trip to work by a given mode, after that heads to a secondary activity and then back home. The spatial path flow includes information about times spent at the primary and secondary activity, the total daily travel time and chosen locations and modes.

In summary, the presented methodology offers a possibility to estimate spatial path flows considering constraints of time and space. Naturally, as models are simplifications of what is observed in real life, the approach is based on several assumptions. Nonetheless, it provides a hierarchical and applicable framework for solving the problem of modeling a complex tour of activity and travel patterns. So far, the explanations remained generic not taking into account real data. In consequence, the task remains for the upcoming chapters a) to introduce the necessary information and results of the statistical analyzes to provide the (threshold) values of constraints and b) to develop and apply an algorithm allowing for the estimation of spatial path flows. A more elaborated discussion on the advantages and disadvantages associated with the approach is subject of the final chapter.

The following Figure 3.2 summarizes the methodology and indicates calculation steps and required information. In addition to the generic descriptions given so far, the flow chart already makes reference to data and model output used for the empirical application. All sources as well as preparation and analysis of the respective data are described in the upcoming chapters. The arrows reflect the sequential procedure as well as the points at which interactions between steps are required.

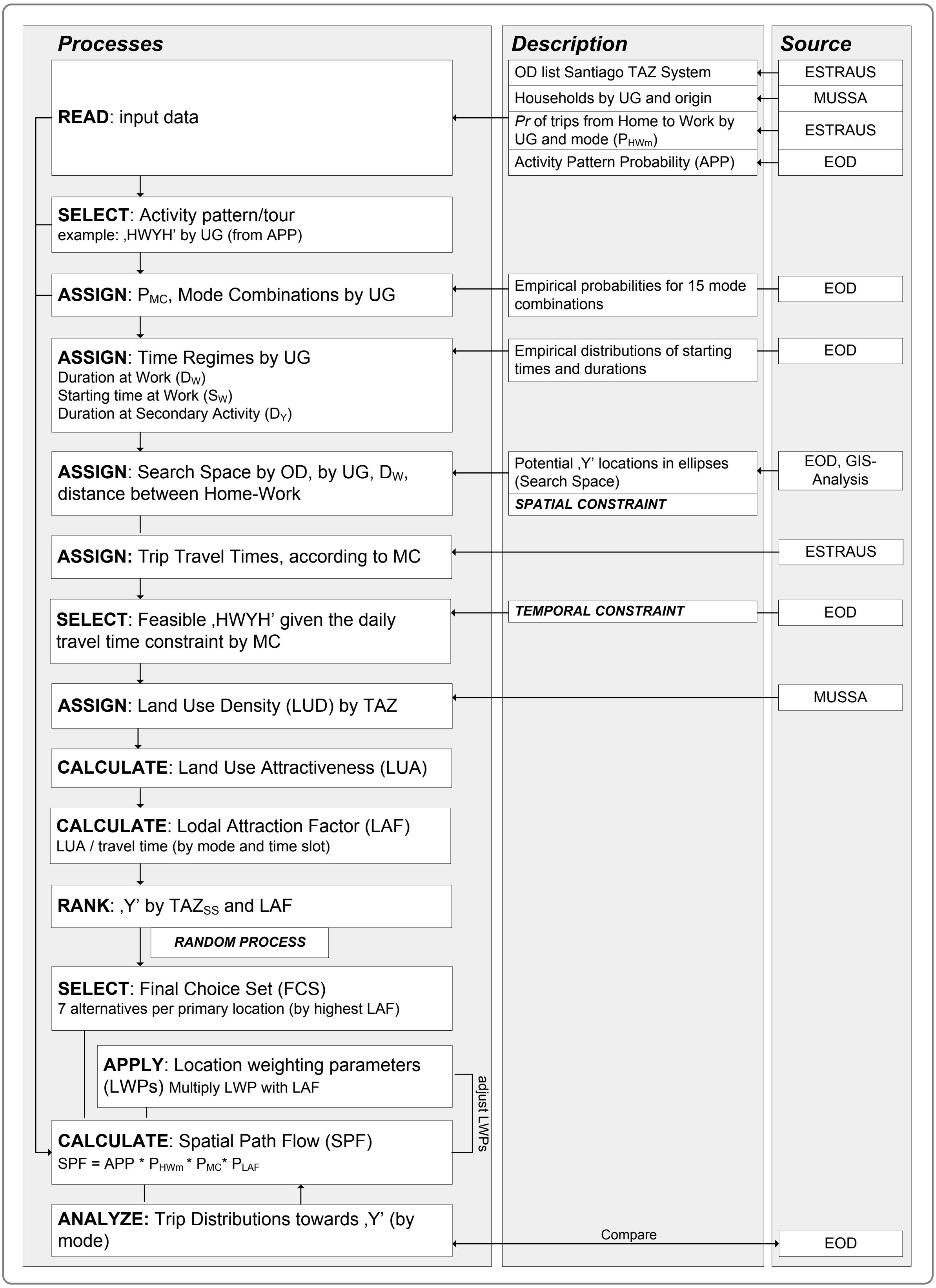


Figure 3.2: Methodology – calculation procedure

3.3 Research questions

The empirical implementation strictly follows the methodology described above. However, more concrete questions arise as soon as it comes to the actual implementation of the methodology with the use of empirical data. Some data requirements already became evident; nonetheless, specific values of the model parameters have not yet been defined. Respectively, the more specific research questions address these issues of values and parameters, all of them being derived from the central research question formulated as follows:

What are appropriate time and space constraints (and how should values for them look like) that allow for building a probabilistic tour-based model of activity and travel patterns?

The subordinated research questions related to the central research question can be summarized as follows:

What influence does the duration of the primary activity have on the search space?

According to the conceptual characteristics of a hierarchical choice-making process, it is assumed that time spent on primary activity influences the adjacent decisions for times and locations. For example, a longer stay at the workplace reduces the time left for any shopping or leisure activities and probably shortens the size of the spatial area where this activity is realized. However, assuming this influence as of relevance, how strong is this influence and how could it be practically asserted in the probabilistic model?

What are the elements that determine detours from the axis of primary locations and what are appropriate empirical values to describe detours?

Given a limited daily travel time and the efforts involved to reach the locations, it is assumed that there exists some restricted geographical search space containing most of the potential travel options. It remains still to define empirically whether this space should be differentiated by some attribute inherent to the model or what suitable thresholds for the detour factors would be. Eventually, the research question is also related to the issue of how ‘well’ – in comparison to empirical observations – the detour factor defines the space in which decisions for secondary activities locations are taken.

How can daily travel times for an entire tour be considered, and to what extent do they constrain the search space?

Intuitively, we suggest that travel time varies according to the mode combinations used for a tour. For instance, it is expected that non-motorized mode combinations, especially walking tours, will show smaller daily travel times than motorized combinations, for instance, a tour by car. The remaining questions are on the variations of daily travel times according to the mode combination: what combinations should actually be considered, how could time thresholds be practically included into the approach and what is its effect on the search space and, respectively, on the number of spatial options?

How can the location choice for a secondary activity be adjusted in such way that the observed choices are reproduced in the best way possible?

This question is very much related to the common challenge of calibrating the location choice. This might be done in a similar way as the established transport models do, i.e. through an iterative process of parameter search. But the question remains: how these parameters can be integrated and how they best fit into the calculation approach.

Generally, to what extent is the approach able to reproduce the variations of travel behavior?

The question is very much related to the general concern that a model is only capable to reproduce behavioral realism in compliance with the implied simplifications. This concern is of relevance here as the approach is probabilistic by design and aims at the reduction of potential location choices. Consequently, the approach on purpose excludes choice options of little relevance. The objective is to describe and order the most probable options for a location choice rather than for randomly occurring events. The same applies when it comes to the definition of relevant mode combinations to consider, as only the most relevant combinations should be included. Hence, the questions remain open, how can 'relevance' be defined and how much of the behavioral variations can be reproduced by the approach.

The questions raised are addressed in the following chapters by a profound analysis of empirical data. Basically, the attempt is to respond to the research questions and at the same time replicate in a best possible way the travel behavior observed in Santiago's travel survey. As a 'manual' for the model development serves the hierarchical framework and the step-by-step methodology recently introduced.

4 Data and models: evidences from Santiago

The implementation of the theoretic framework is based on several data sources, particularly, on a large-scale Travel and Household Survey and on results obtained from the applications of Santiago land-use and transport models. These sources provide important input to the method and require in particular in the case of the travel survey substantial pre-processing efforts. Given that the empirical application is based on this information and as we will frequently refer to these sources; at this point, we introduce the main data and model sources in brief. First of all, the given geography and its disaggregation into smaller units, i.e. Traffic Analysis Zones (TAZ), are illustrated. The spatial levels considered comprise a macro scale, represented by Santiago Traffic Analysis Zones, and a micro level reflected at the residential block level (see Figure 4.1).

Figure 4.1: Maps of spatial levels in Santiago

Annotation: left figure: 618 TAZ (thin grey border lines) and borders of the 38 municipalities (black border lines) pertaining to the Metropolitan Area of Santiago; in addition, the 6 even larger city sectors are included (thick grey lines); right figure: municipality of Santiago-Centro (thick grey border line), TAZ (grey border lines), and about 50.000 residential blocks (thin grey border lines)

The 618 TAZ build the background of the established aggregated transport model of the city of Santiago and were also used for the estimation of spatial paths. The information about the land-use, provided by the city's land-use model, is available at the TAZ level, too. The information about activity and travel patterns, for instance, about distances between activities, i.e. trip lengths, was collected on Santiago's residential block level.

4.1 Santiago transport and land-use models

There exist a long tradition and vast experience in the application of models of transport and land-use in urban planning in Santiago. First applications of the trip-based transport model ESTRAUS date back to the development of the four-step model in the 1980s (Mideplan, 2005; Florian et al., 2002; de Cea et al., 2003). The development of the land-use model

MUSSA and of the transport model ESTRAUS was financed by the Chilean government. Today, planning entities make use of these models to analyze infrastructure projects or to make predictions for the city's land market (Martínez, 1996; Martínez, 2008; Sectra, 2008). After successful applications of the land-use model MUSSA for Santiago, the model has been recently integrated into a commercial modeling suite, which is applied in various cities worldwide (Citilabs, 2010).

MUSSA is software and refers to a mathematical model designed to describe, predict, simulate and analyze the urban real estate market and allow planners to forecast and simulate its economic equilibrium under various demographic, macroeconomic and regulatory scenarios. The user simulates the real estate market for these scenarios and assesses the economic impact on the city. According to the model, urban real estate market consumers, whether in households or in their economic activities, are located at places where they are the highest bidders in the market, with real estate assigned to a particular use according to 'auction' rules. This approach sees the rent of each property defined by the highest bid submitted. The socio-economic household attributes are included here in order to consider household life cycles, while zone quality indices, for instance, reflect accessibility to the transport system. Urban management policies designed to stimulate (subsidies) or restrict (taxes) the location of activities in the city can be added to the analysis with MUSSA.

The modeling outputs of MUSSA in form of a detailed description of the land-use of the city are important for this work. The model predicts the location of households by 13 types and of firms by 5 economic activities at different geographical levels, whereas the most disaggregated level corresponds to Santiago TAZ. MUSSA is seen as an input to the ESTRAUS urban transport model. Localizing household user groups differentiated by car ownership and income class is essential for calculating trip generation and attraction of each zone of the transport system. Likewise, the transport model passes to MUSSA information relating to spatial measures dealing with accessibility and the attractiveness of zones. Consumers (household user groups or firms) regard these access measures as variables in the process of choosing a location. Strategic projects with impact on the transport system are considered in the land-use system, i.e. the localization of user groups and firms adjusted with MUSSA in turn affects the transport demand calculated by ESTRAUS (Sectra, 2008).

ESTRAUS is a classic four-step transport demand model that estimates trip generation (number of trips by user group and zone), performs trip distribution according to the attractiveness of each zone in the area, and partitions modes by dividing calculated traffic flows by the modes in question. In the fourth step, the estimated demand is iteratively assigned to public and private transport networks via Origin Destination Matrices until equilibrium between supply (networks) and demand (users) is achieved. ESTRAUS is trip-based, and its current version predicts city transport demand disaggregated by the 13 household user groups, three travel purposes (work, education, and other) and for two temporal periods (morning and afternoon rush hours). Figure 4.2 outlines the components of MUSSA and ESTRAUS and their interaction.

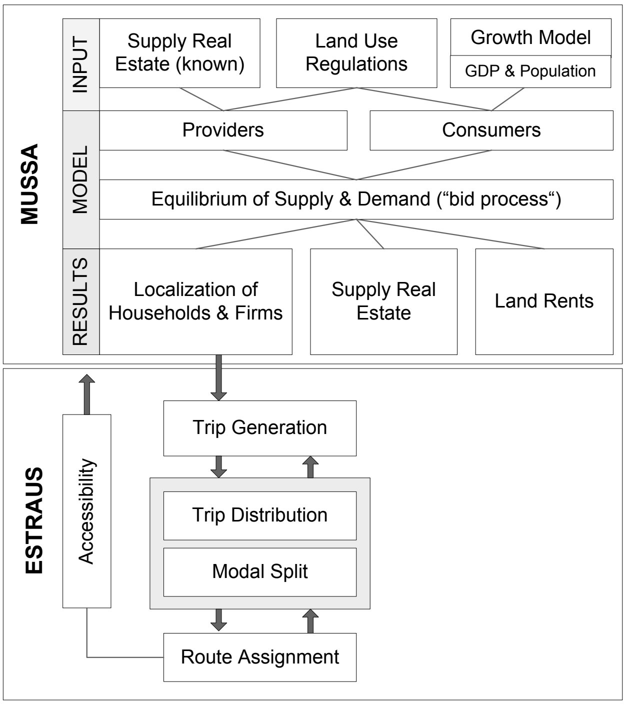


Figure 4.2: Interaction between land-use and transport models

Annotation: The Growth Model indicated above is externally applied to predict the overall number of household user groups that are located via MUSSA. It is principally based on a regression analysis of historical information from different countries on the interaction of GDP, demography and motorization.

The output of both models is manifold information about the transport and land-use system performance. In general, this includes traffic volumes on links, congestion levels or land rents and localized firms and households. The following Figure 4.3 summarizes the main information required and provided by runs of MUSSA and ESTRAUS.

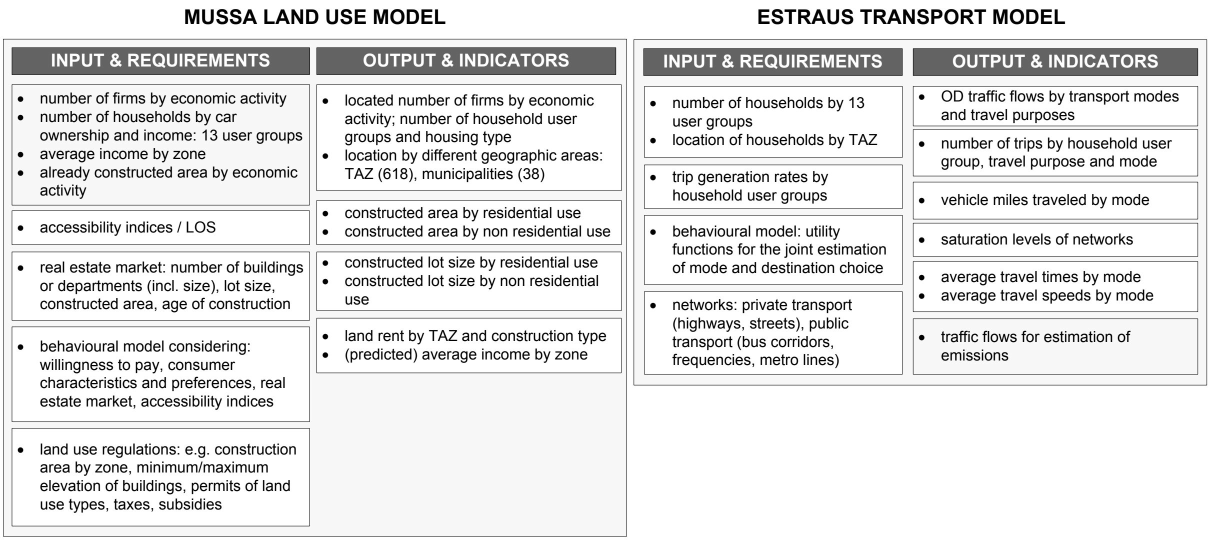


Figure 4.3: Input and output of Santiago's land-use and transport models

For this work, some selected outputs of the models are further processed and integrated into the calculation approach. Basically and according to the figure above, the following information is included:

- ESTRAUS: OD traffic flows, i.e. commuting matrices by household user group and mode (morning peak hour);
- ESTRAUS: Accessibility indices via level-of-service matrices by mode (morning peak hour);
- MUSSA: Built area in square meters by non-residential use, specifically of commerce and service-based activities by TAZ;
- MUSSA: Number of households by user group and TAZ.

The commuting matrices determine the relation between the home and work locations; the level-of-service matrices describe the network supply and thus its differences in accessibility throughout the city. The commerce and service-related land-use is included to describe spatial attractiveness of TAZ with regard to activities of discretionary character, i.e. leisure or shopping trips. The adoption of commuting matrices comes along with the adaptation of the underlying classification of the demand into 13 household user groups. In the following chapters, we will refer to these user groups because they are utilized as distinguishing feature in the data analysis, and, finally, we will estimate spatial path flows for each of these groups. The following Table 4.1 shows the differentiation of user groups according to household income and number of vehicles available.

Table 4.1: Household user groups by income and car ownership

Disaggregated groups	Grouped by income	Grouped by car ownership	N Cars	Income in Chilean Pesos (CHP)	Abbreviation
1	1	1	0	<= 148.226 CHP	**<= 150**
2		2	1+		
3	2	1	0	148.227 to 296.452 CHP	**> 150 & <= 300**
4		2	1+		
5	3	1	0	296.453 to 592.904 CHP	**> 300 & <= 600**
6		2	1		
7			2+		
8	4	1	0	592.905 to 1.185.808 CHP	**> 600 & <= 1.200**
9		2	1		
10			2+		
11	5	1	0	> 1.185.809 CHP	**> 1.200**
12		2	1		
13			2+		

Annotation: Abbreviations as shown in the right column are used throughout the thesis to improve readability.

The information provided by MUSSA goes back to a model run of 2002 and, accordingly, represents land-use distribution of this year. This coincides with the Travel and Household Survey carried out in 2001. The ESTRAUS information regarding commuting matrices and level-of-service matrices dates back to 2007. With the introduction of the new public transport system Transantiago in 2007, some structural model elements of ESTRAUS were adapted. In particular, public transport modes (bus, metro), which earlier had been treated individually, were included in an integrated way in 2007, thereby reflecting the integration of both modes. In addition, between 2001 and 2007 the spatial zoning system was refined, i.e. the number of

TAZ was increased. Hence, it was decided to include the newer zoning system and to use the more recent accessibility measures.

4.2 EOD: Santiago Travel and Household Survey

The other main source for the description of activity and travel behavior in Santiago is the Travel and Household Survey 'Encuesta Origen-Destino 2001', abbreviated EOD (Sectra, 2002; Dictuc, 2001). The main objectives of the EOD were to provide a detailed description of the city travel patterns that would allow the estimation of origin-destination matrices and to generate necessary parameters for the urban transport model ESTRAUS. Beside the traditional household survey about individual travel behavior on a single day, the EOD also included surveys at transfer points between modes, traffic counting at cordons and throughout the city, as well as collection of data about tariffs and level-of-services. The information was collected for all days of the week (Monday to Sunday). The EOD sample provides information about approximately 150,000 trips as well as socio-demographic and economic characteristics of nearly 60,000 people living in 15,000 households. Respectively, three data sets about trips, persons and households are available. Furthermore, the survey collected geographical coordinates for all observed trips on residential block level. The following Figure 4.4 depicts the geographical solution of the information available in the EOD.

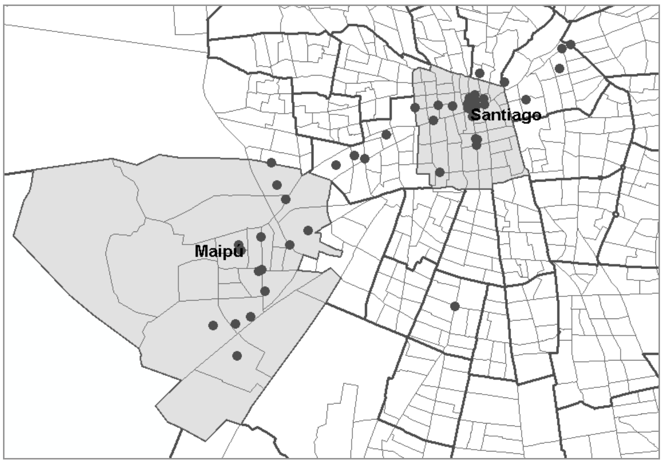


Figure 4.4: Example of geocoded EOD survey information

Annotation: The grey points reflect on residential block level all activities visited beside the Home and Work locations of all surveyed individuals living either in the municipality of Santiago or in Maipú.

4.3 Activity and travel behavior in Santiago

The empirical information about activity and travel behavior that is discussed in following sections is based on the analysis of the EOD. The data set was used to generate comprehensive activity patterns comprising the information about all activities and trips one person conducts during one day. The information available includes type of activity, starting time and duration of activities and trips, duration of at-home and out-of-home stay, used modes and associated person- and household-related variables like age, gender, household income or number of vehicles per household. The following sections concentrate on the number of operations indispensable for providing of a proper data set for the analysis. First, a short summary is given upon the main data preparation steps that lead to a reduced but perfectly suited data set for the subsequent analysis. After that, based on the reduced data set, approaches for unambiguous definition of activities of primary and secondary priority are described. The final section of chapter 0 provides a comprehensive typology of activity pattern types. This typology then serves as a foundation for the identification of tours and activity pattern and their respective analysis in chapter 0.

4.3.1 Summary on main data preparation steps

The data provided by the EOD may be best introduced by demonstrating what an example of a single person reveals in the survey. Respectively, Figure 4.5 illustrates the information provided in the trip data set including reported activities, trips, timing aspects and modes used.

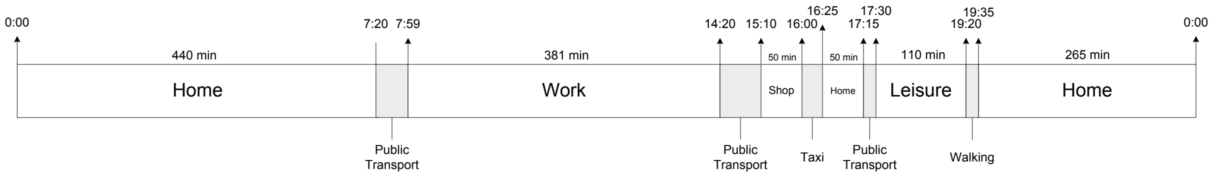


Figure 4.5: Example of an EOD daily activity pattern

This example is real and is reported in the data set. The grey parts represent trips in between activities and the associated modes used. The width of each segment reflects the share of time consumed for an activity or for a trip in the course of one day (24h, 1,440 minutes). In this example, the person undertook three out-of-home activities with one intermediate stop at home, spending altogether 755 minutes at home, 541 minutes on activities and 144 minutes traveling. We can provide additional information on the person and household characteristics: this is a 61-year-old woman; she is employed full-time, she has no driving license, she lives in a multi-person household without children; there is no car in the household and she lives in the municipality of La Florida.

The provision of a comprehensive EOD data set for the generation of activity pattern and tours required substantial work on data preparation, cleaning and testing, since data analysis based on entire tours or activity pattern requires a proper listing of all activities and trips, including all attributes. For instance, if one element (activity site visited or mode used) of the tour is missing, this tour is excluded from the analysis, whereas it would still be usable for a trip-based analysis (Davidson et al., 2007, p. 475). Furthermore, it should be recognized that "conventional household interview surveys are characterized by a significant percentage of non-closed tours and daily patterns with 'gaps' where return-home trips are frequently missing" (ibid.). Newsome et al. (1998) conclude that "it is sometimes necessary to use travel diary data as a surrogate (…), but it requires a great deal of data processing to put them in a form useful for examining activities" (p. 361).

However, not every single data processing step conducted in the EOD is subject of discussion. Instead, we make reference to the respective program files that describe the modification of the raw data set in detail (see Digital Annex, Annex-Table 6: Overview of SPSS-Syntax used for EOD data mining). The creation of particular new variables is described there in more detail, too. As a result of the data processing the sample was slightly reduced. The following Figure 4.6 summarizes the reduction steps and outlines the underlying reasons.

	Original data set: 46.168 activity pattern; equal to 153.413 trips	**100%**
1	**Remaining: 45.992 activity pattern // Reason: incomplete starting and ending times of trips**	**99.6%**
2	**45.102 activity pattern // pattern with more than 7 activities, i.e. 8 trips**	**97.7%**
3	**44.425 activity pattern // consecutive ‚Home' activities**	**96.2%**
4	**44.235 activity pattern // consecutive ‚Work' or ‚Education' activities**	**93.7%**
5	**41.970 activity pattern // incomplete pattern; e.g. missing ‚Home' activity at the end**	**90.9%**
6	**41.922 activity pattern // error in summed activity and travel time (> 24h)**	**90.8%**

Figure 4.6: Data mining: reduction steps

The reduction steps require some further explanations:

1. Incomplete information about temporal aspects (starting or ending times) of one trip led to the elimination of all trips conducted by the respective person, as no coherent pattern could be build (0.4%).
2. Activity patterns with more than seven activities were eliminated to reduce complexity; the threshold of seven activities (eight trips) appeared appropriate as only 1.9% of all EOD activity pattern contained between eight and a maximum of 23 activities.
3. Activity pattern containing two consecutive 'Home' activities were interpreted as incorrect answer and were removed (1.5%).
4. The same applied to consecutively reported 'Work' or 'Education' activities as it was assumed that no trips can occur between the same locations (2.5%).
5. Unclosed activity patterns, meaning patterns that either did not start or did not end at home, were removed (2.8%).
6. Activity pattern with unreasonable summed times of activities and trips (> 24h) were removed (0.1%).

It is important to discuss briefly the issue that the modification of the raw data leads to a reduction of the data set for analysis. The reduction of the original sample size can be problematic if too many cases are eliminated and representativeness is jeopardized. Another problem emerges if the elimination of cases does not occur systematically over the entire sample. Given the fact that the EOD data set is a representative sample of Santiago's population, a non-systematic elimination of cases would lead to a biased sample, eventually reducing the explanatory power of the data. Hence, after the reduction we tested whether the data set remained representative in comparison to the initial sample. This was done grouping the data set into the 13 household user groups as these groups were used later as socio-economic categories throughout the statistical analysis. Figure 4.7 compares the shares of household user groups in the number of activity pattern before and after the application of the reduction rules.

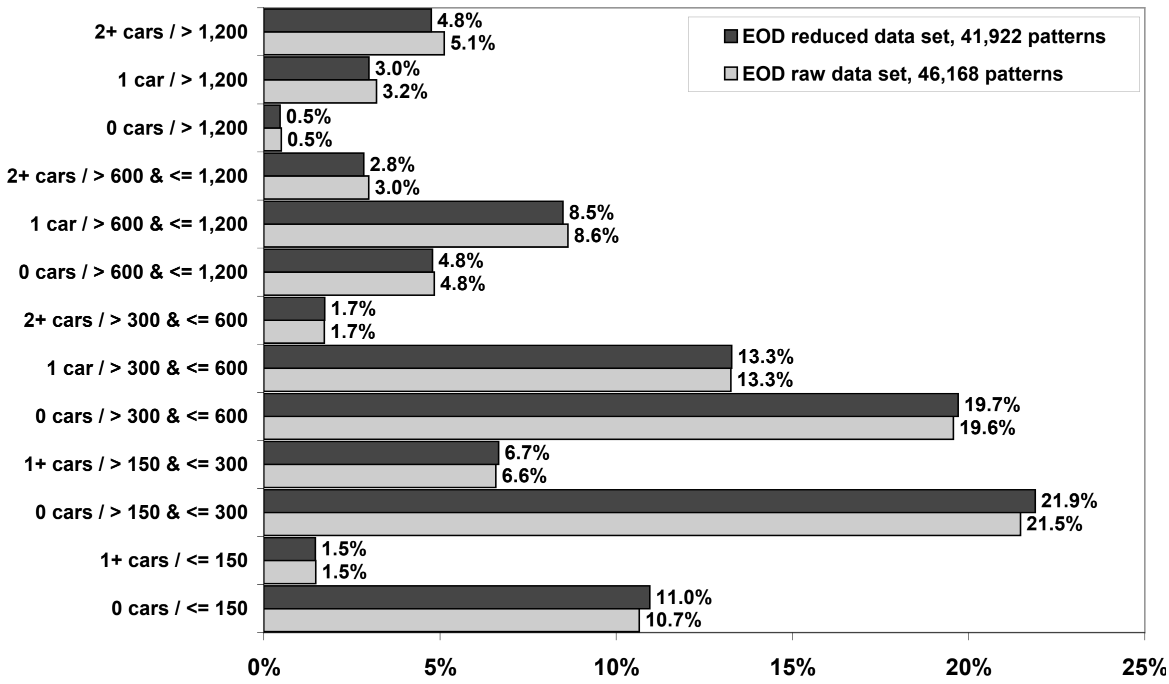


Figure 4.7: Comparison of EOD data set - before and after data mining

In sum, the differences vary between -0.4% (user group: 0 cars / > 150 & <= 300) and +0.3% (user group: 2+ cars / > 1.200) resulting in an overall standard deviation of 0.19. Based on these rather small deviations it is concluded that the reduced data set provides a level of representativeness comparable to the one of the original sample. In consequence, any following data analysis is based on the reduced data set as described in this section.

4.3.2 Activity prioritization

In chapter 2 and in the discussion on different travel demand model types it became evident, that the definition of priorities between activities plays an important role when it comes to the modeling of tours and of activity patterns. We introduced the concept of macro and micro scales applied to the type of activity, distinguishing between activities of primary and secondary importance. By that we assume that decisions associated, for instance, to a recreational activity like visiting a gym (time, mode, location) will depend on the decisions taken on a more important activity, e.g., work or education. This is in line with other approaches where mandatory activities are assumed to be fixed while discretionary activities are more flexible in time and space (Doherty, 2006, p. 518; Doherty and Mohammadian, 2011, p. 45). However, the assumed hierarchy of activities priorities is disputable and recent research reveals that the order in which activities are actually planned does not follow any continuum of activity type importance. "(…) activities are not planned according to a fixed hierarchy of activity types, rather the activity characteristics play more important role in identifying the order of that activity in the tour" (Doherty and Mohammadian, 2011, p. 57). The same authors come to the conclusion that activity duration mainly serves to rank activities, where longer lasting activities tend to be planned earlier (ibid., p. 60). Generally, the interest in the order in which activities are actually planned is related to models that incorporate planning and scheduling of activities (Buliung and Kanaroglou, 2007, p. 155). This is different to our approach, where the interest upon activities importance is related to the question of hierarchies in decision-making and whether there exist certain activities that constrain decisions (times, modes, location) related to subsequent activities. Respectively, at this point we have to make a distinction between the order in which activities are planned (adjusted, removed) and the order in which they should be (spatially) modeled. Even though a discretionary activity of very long duration may be planned first, the spatially fixed activities/locations, such as the home, work or education locations still have an influence on the spatial interrelations of activities.

Travel survey data often provides no information regarding priorities between activities conducted during a day (or week). Given the importance of activity prioritization modeling, it is surprising that surveys are carried out with no regard to this issue. For instance, Bowman and Ben-Akiva (2000, p. 9) recommend that if a model is based on activity priority, the respective information should be collected directly in the activity and travel surveys. Buliung (2005, p. 27) argues the same: "One reason for our current lack of knowledge is that respondents have rarely been asked to record information concerning activity prioritization when completing activity-travel-surveys." The same applies for the EOD data, making it impossible to retrace the order in which activities were planned. Nor it is asked if some activity sites – beside the home, work and education locations – are visited frequently. This information would allow assigning higher priorities to these activities over others. Nonetheless, it is possible to structure activities ex-post depending on their assumed importance. Hence, the objective of this section is to test our theoretical assumption of existing macro and micro scales against the application of a set of classification approaches discussed in literature. In view of that, we assign one single primary activity to a pattern while all other activities are denominated as of secondary importance. The intention we pursue here is twofold: we want to find out whether our a priori denomination of mandatory activities (work, education) remains valid if we classify activities only by their duration, i.e. by the time an individual spent (invested). Based on that we aim at deciding upon an approach that assigns priorities on pattern also in cases where no work or education activities are reported. Notice that at this point we assign one primary activity per pattern, not per tour. This is acceptable regarding the analytic question followed here but may be expanded assigning one primary activity per tour once several tours in a row are modeled.

It was argued that the priority between activities is most generally characterized by the fact that compulsive activities receive higher priorities than discretionary activities. Work and educational (school, university) purposes are mostly defined as those of primary importance; other activities like shopping or leisure are classified as of secondary importance (Doherty, 2006, p. 518). Cascetta (2009, p. 221) extends this two-level classification proposing a three-level hierarchy among activities, with the first level of work, study and school activities; a second level comprising professional, business or service activities; and a third level of any other purpose. If a model first identifies a primary activity and then arranges subsequent decisions around the primary activity, even in case of tours without a work or education activity we must define unambiguously which activity represents the (spatial) anchor point. Even if Cascetta's three-level hierarchy is applied, activities of the same hierarchical level may appear within one tour. In these cases, he suggests that the longest activity (in duration) is the one of the highest priority. If then activities appear equal, the activity in maximum distance is defined as of primary importance (ibid, p. 221). In a similar manner, Vovsha et al. (2004, p. 13) depict a three-level hierarchy of activities: first, mandatory activities (work, university, school); second, maintenance activities (shopping, banking, seeing doctor); and third, discretionary activities (social and recreational activities, eating out). Notice that any of these – intuitively understandable – attempts are likely to produce biased results. Actually, there seems to be no other option than actually asking respondents during the survey if certain activities were more important (or more frequently visited) than others. However, as Vovsha takes an approach similar to Cascetta's but suggests a more precise definition of activities of mandatory, maintenance and discretionary character; we consider his classification for the subsequent analysis, too. Hence, in the following we summarize three classification options and apply the respective rules to the EOD data for the definition of primary activities in activity pattern. Our primary interest is on proving if our theoretical argument of macro and micro scales holds where we argued that hierarchical scales are related to the amount of resources (time) spent. With regard to the empirical analysis the question then is whether the definition of primary activities following our 'resource argument' – applied in form of activity duration – matches with the conceptual approaches provided by the cited authors.

In the case of the EOD, in total 12 travel purposes can be assigned to hierarchical levels of activities: Work, Education, Business, Shopping, Visit, Services, Leisure, Health, Bring Somebody, Bring Something, Eat and Other. In accordance with the remarks above, we test the following assignment rules:

- Two-level hierarchy: work and education are defined a priori as primary activities (if both activities occur, the longer in duration activity is considered the primary activity); if no work or education activities are reported, the longest in duration activity is defined as primary activity.
- Three-level hierarchy: work and education are defined as primary activities; if no work or education activities are reported, 'health', 'bring somebody', 'bring something', 'shopping' and 'services' are defined as primary activities (if more than one of them occur, the one of longest duration is the primary activity); if none of the previous activities is reported, the primary activity is 'visit', 'leisure', 'eat' or 'other' (again, ordered by duration if several of them occur).
- Activity duration: the longest in duration activity is defined as primary activity.

Regarding the two-level hierarchy the following distributions for workdays and for a weekend are obtained (see Figure 4.8). Pictured are the shares of activity pattern differentiated by the primary activity they contain; defined according to the rules of the two-level hierarchy (see above).

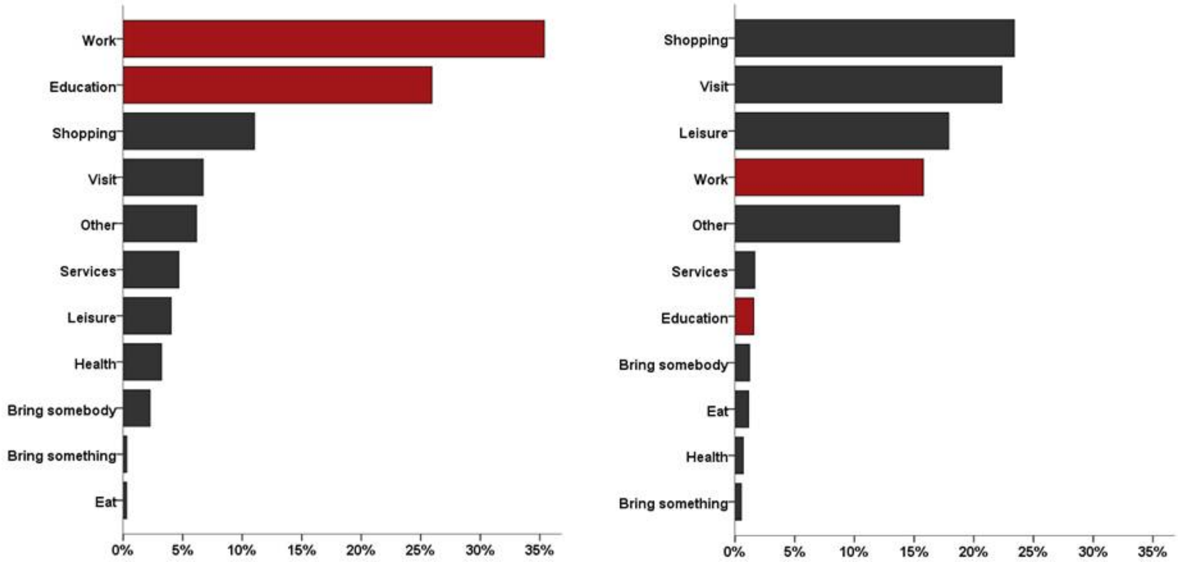


Figure 4.8: Activity prioritization: two-level hierarchy

Annotation: left: Monday-Friday / (N=31,992); right: Saturday-Sunday / (N=10,348)

In case of workdays, about 61% of all patterns include work or education activities which are defined as of primary importance. The remaining 39% are ranked in respect to the longest activity within each pattern. There is no clear dominance among the remaining travel purposes; the ranking order presented here follows the general number of observations for 'Shopping', 'Visit' and 'Other' (see also Annex-Table 1, p. 145). The number of observations in the EOD influences the order by duration; for instance, if 'Shopping' activities are responsible for 21.9% of all trips/activities (see Annex-Table 1), it is likely that more patterns are dominated by 'Shopping' activity. It is interesting to look closer into the travel purposes where the position according to the duration does not correspond to the order given by the number of observations in the data set. This refers to 'Services' and 'Bring Somebody' activities. If a 'Services' activity appears, it is often the most important activity in the pattern, whereas classification by duration indicates position four among the nine discretionary activities. Regarding the number of observations indicated in Annex-Table 1, the 'Services' activity is ranked on position six. The opposite can be observed for the 'Bring Somebody' activity. In the EOD this activity is responsible for about 6.9% of all trips/activities and is on position five among the discretionary activities. Regarding the ranking by duration, the activity plays a less important role and is ranked on position seven. This indicates that the activity often takes less time and thus is less often the most important activity within a pattern.

For the weekend, the activities are arranged in a completely different order, which in principle accomplishes the expectations. The dominant activities are those of 'Shopping' and 'Visit'. 'Work' and 'Education' are not dominant any more, although about 16% of all patterns still include a work activity, most probably due to working activities on Saturdays. The ranking order among all discretionary activities is slightly different from that of the workday. 'Leisure' activities are now more often the longest activity, thus, the primary activity within the pattern. 'Services' and 'Health' activities are the longest activity less often. This is most likely due to the fact that often they cannot be conducted at the weekend, as both activities partially depend on opening hours of respective facilities. Complementary to Figure 4.8, the following Table 4.2 shows the mean and median for durations of the primary activities.

Table 4.2: Primary activities durations

	workday (Monday-Friday)					weekend (Saturday, Sunday)				
	Primary activity duration									
	Valid N	Valid N in %	Median	Mean	Standard Deviation	Valid N	Valid N in %	Median	Mean	Standard Deviation
Work	11,316	35.4	565	508	191	1,632	15.8	475	461	207
Education	8,296	25.9	335	354	118	163	1.6	235	264	139
Shopping	3,531	11.0	40	58	69	2,420	23.4	60	74	66
Visit	2,151	6.7	210	251	175	2,313	22.4	245	283	167
Other	1,972	6.2	115	181	183	1,426	13.8	133	180	158
Services	1,497	4.7	80	108	98	172	1.7	100	146	137
Leisure	1,290	4.0	150	187	139	1,852	17.9	180	209	138
Health	1,028	3.2	100	129	99	71	0.7	100	128	105
Bring somebody	723	2.3	10	50	108	127	1.2	45	98	126
Bring something	95	0.3	61	103	125	54	0.5	40	89	112
Eat	93	0.3	115	148	103	118	1.1	152	178	97

Annotation: left: workday / (N=31,992); right: weekend / (N=10,348); ordered by column 'Number' (workday)

In sum, the analysis of primary activities and discretionary activities indicates that to a certain extent the order reflects the share the trips represent in the EOD. More frequently reported activities are also likely to be more often the longest in duration. This is especially the case for those patterns where only one activity is conducted. In these cases, the single activity is automatically the longest, thus, the primary activity. Deviances from the ordering scheme have been detected for 'Services' and 'Bring Somebody' activities for workdays and for the 'Leisure', 'Services' and 'Health' activities at the weekend.

In the following Figure 4.9 again the shares of activity pattern differentiated by the primary activity they contain are pictured; now defined according to the rules of the three-level hierarchy (see rules above).

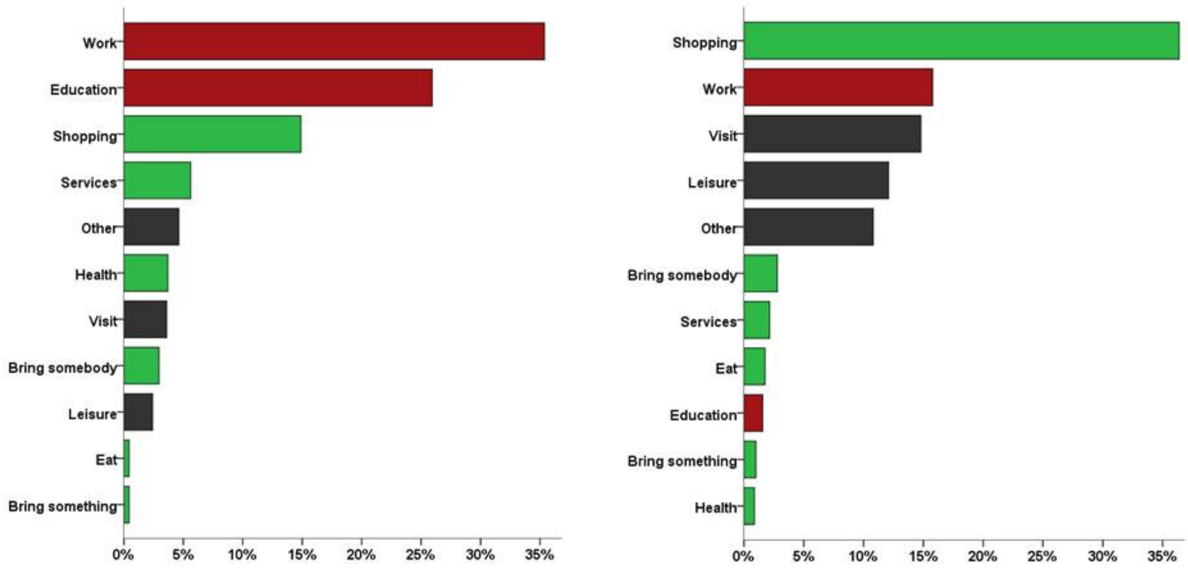


Figure 4.9: Activity prioritization: three-level hierarchy

Annotation: left: Monday-Friday / (N=31,992); right: Saturday-Sunday / (N=10,348); brown: first priority, green: second priority, grey: third priority

First of all, for workday and weekend the shares of primary activities covered by work and education activities remain the same between Figure 4.8 and Figure 4.9 due to the application of the same rule. However, arrangement of all other activities indicates some interesting differences. The three-level hierarchical approach shows higher shares for 'Shopping' and 'Services' and smaller shares for 'Visit', 'Other' and 'Leisure' activities in comparison to the two-level approach (Figure 4.8). More often an activity was defined as of primary importance, despite relatively short activity duration. Essentially, the three-level hierarchy process favors shorter activities in activity pattern to become the primary activity. This aspect gets even more evident comparing the results between the weekends. In this case, 'Shopping' is by far most often denominated as of primary importance while in the two-level approach the distribution among the activities of 'Shopping', 'Work', 'Visit', 'Leisure' and 'Other' is more even. We might conclude that a higher share of primary activities in pattern with a 'Shopping' activity appears at cost of the third-order discretionary activities of 'Visit', 'Leisure' and 'Other'.

Finally, Figure 4.10 shows the shares of activity pattern differentiated by the primary activity they contain according to the activity duration (see rules above).

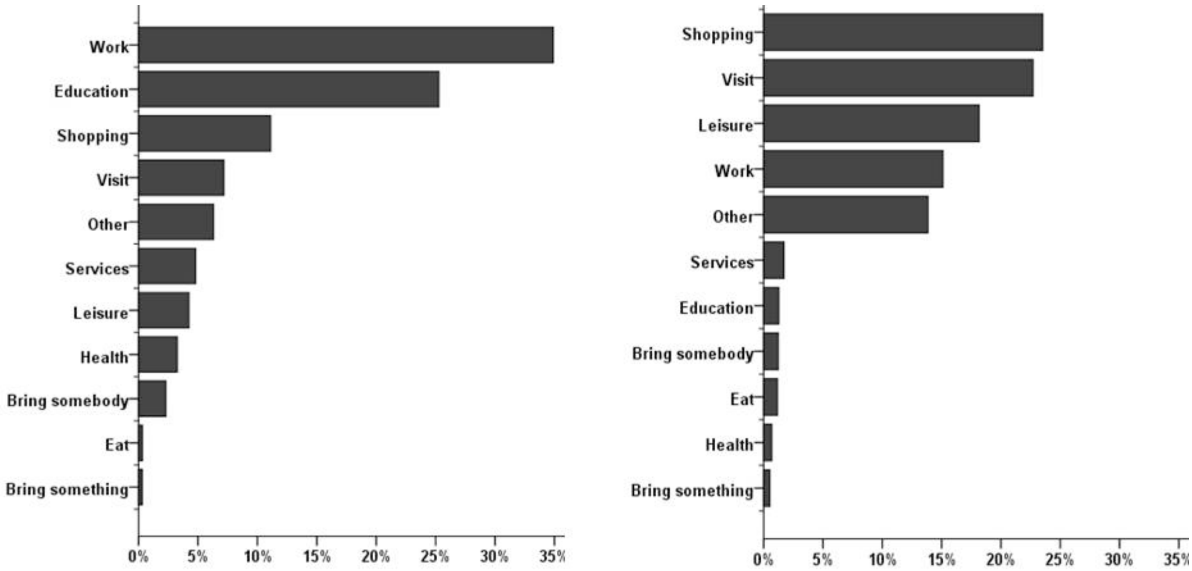


Figure 4.10: Activity prioritization: activity duration

Annotation: left: Monday-Friday / (N=31,992); right: Saturday-Sunday (N=10,348)

The differences between the two-level hierarchy and the solely prioritization by activity duration are minor. This is interesting as in the former approach work and education activities were a priori defined as primary activities. However, this final prioritization reveals that if a work or education activity is conducted, it nearly always is the longest activity in duration. In only 1.1% of the observations (referring to the Monday-through-Friday results) this was not the case, meaning that other activities of the pattern took longer than the reported work or education activity. Summing up, the two definitions of the two-level hierarchy and the prioritization by activity duration are nearly equivalent with the important difference that the former follows an arbitrary definition while the latter is objectively defined applying the criterion of time consumption. We conclude that in this case our initial assumption about the existence of macro and micro scales basing on the 'resource argument' is confirmed. This has the effect that both the two-level hierarchy and the activity-duration-based approach are similar regarding the assignation of activity importance.

As long as survey data does not indicate which of the activities conducted during the day was of higher or of the highest importance, alternative rules are required as applied here. Given our assumption that the relevance one assigns to a decision is related to the required investment (in this case, time), the approach of a two-level hierarchy seems to be a suitable solution. The analysis confirmed the relevance of work and education activities in case of Santiago. The relevance is due to their fixed locations and thus, their influence on spatially dependent decisions of location choice. Their primary status was also confirmed with regard to the time an individual spent on these activities. In consequence, we make use of the two-level hierarchy when it comes to the creation of a pattern and tour typology in the section to follow (see 4.3.3). Notice that at later stages of this work we limit the analysis on the use case of a tour of type 'HXYH'. However, the assignment of primary and secondary activities importance already arranges the entire data in such way that an expansion of the methodology to other tours/pattern becomes possible.

Using a tour of type 'HXYH' as an example implies that no further differentiation for secondary activity 'Y' is made. However, we recognize from this section that activity 'Y' represents a set of quite different activities such as 'shopping', 'leisure' or 'services'. At an earlier stage of this work it was planned to differentiate 'Y' by some specific secondary activities when it comes to modeling. To this purpose, we then focused the analysis on grouping the secondary activities so they would represent a sufficient amount of observations for empirical analysis. However, because of the need for a reasonable amount of empirical observations, we eventually decided to base the analysis on the even more aggregated activity 'Y' comprising all EOD activities of discretionary type. Therefore, it is not necessary to now explain in detail the prior applied methodology for a grouping of the 12 EOD activities into five activities. Instead we move on to the next section where a typology of activity pattern is introduced based on activities 'X' and 'Y'. Nonetheless, the analysis dealing with the steps undertaken to aggregate EOD travel purposes is moved to the Annex chapter.

4.3.3 Tours and activity patterns

In the previous section we focused on single activities and on their priority among all activities of a tour/pattern. Now, we expand the perspective to the entire sequence of activities. Activity patterns and their components, travel tours, are used as a principal unit for the analysis. As we have explained in the introductory chapters, the activity pattern is composed of one or more tours and reflects the activities sequences. At the same time, the consideration of activity patterns and tours implies that the main postulate of activity- and tour-based approaches is met: understanding travel behavior as a demand derived from the interdependent activities. Pattern and tours show different levels of complexity, and their modeling becomes a challenge, especially when several activities are realized without stays at anchor points, for instance, at the home and work locations. Generally, this section has the objective to depict the travel behavior complexity represented by the empirically observed activity patterns of the EOD. Although the model will be employed on the example of a single tour of 'HXYH', the diversity of patterns and tours observed in the EOD help us to conclude on how much of the complexity we might be able to model given the 'lessons learnt' from our use case.

Before proceeding to the discussion of the results, we have to explain how behavioral complexity is actually defined. Complexity refers to the overall number (and their respective share) of different activity pattern types observed in the EOD. The categorization of types depends, for example, on the number of simple and complex tours, meaning the number of outbound activities without getting back home. In addition, complexity refers to the degree to what a pattern is composed of a single or of multiple tours. Also, of the fact whether primary activities of type 'Work' or 'Education' appear or not, serve as a distinguishing feature. In the following, the analysis is reduced to activities and their sequences within patterns. Other dependencies, for instance, among starting times and duration of activities or mode choices are not part of the analysis yet. First, an overall comparison of activity pattern types is provided, defined by:

- The frequency distribution of all EOD activity pattern by aggregated and disaggregated travel purposes.

The distribution of activity pattern is shown in two ways: by differentiating between the five travel purposes (‘Work’ – ‘W’; ‘Education’ – ‘E’; ‘Shopping’ – ‘S’; ‘Leisure’ – ‘L’; ‘Other’ – ‘O’) that resulted from the aggregation process described in Annex-1; and by an even more aggregated approach in which the ‘Work’ and ‘Education’ activities are denoted with ‘X’ and all other activities of ‘Shopping’, ‘Leisure’ and ‘Other’ with ‘Y’; the ‘H’ in each pattern represents the home location. The denomination of ‘X’ and ‘Y’ is helpful for group patterns that show similarities regarding their general division into pattern containing the macro scale activities of ‘Work’ and ‘Education’ or pattern not containing them. It is nonetheless possible to define a primary activity in tours/patterns solely constituted by ‘Y’ activities as it was discussed in the previous section. However, we reduce the analysis at this point to the primary activities of work and education. The use of activity aggregates ‘X’ and ‘Y’ implies that, for instance, the more aggregated pattern of ‘HYHYH’ contains patterns such as ‘HSHLH’ or ‘HOHSH’. Considering this, the following frequency distributions of activity pattern for Santiago as shown in Figure 4.11 were generated:

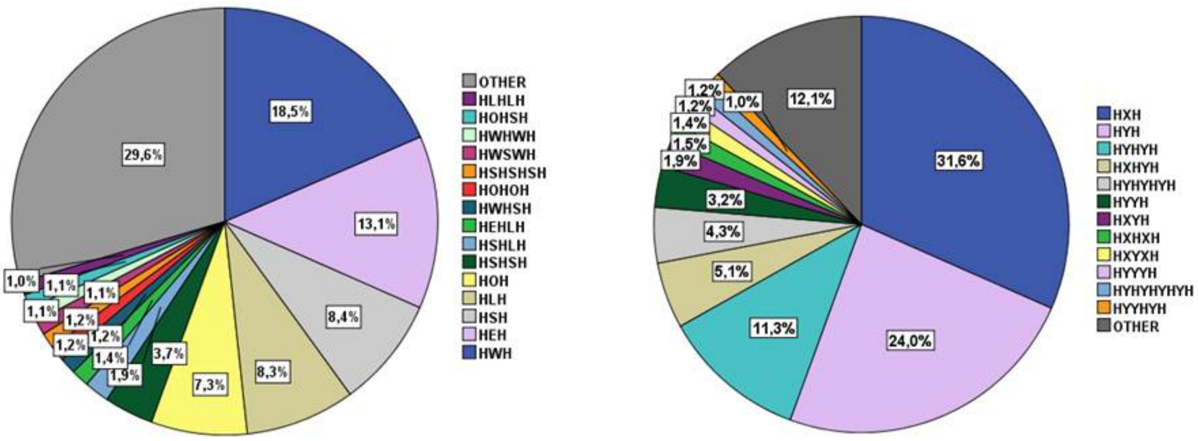


Figure 4.11: Frequency distribution of activity pattern in Santiago

Annotation: left: by five travel purposes; right: by further aggregated travel purposes / (N=41,922)

In case of the five travel purposes in Figure 4.11 (left), 15 out of 1721 patterns represent 70.4% of the entire variation (and 12 out of 292 pattern types in the case of the aggregated purposes, representing 87.9%, figure on the right). Regarding the more aggregated approach the term ‘pattern type’ is used here since type ‘HYH’ comprises patterns of ‘HSH’, ‘HLH’ and ‘HOH’. In both charts the category ‘Other’ contains all activity patterns occurring with a share of less than 1%. In case of the five travel purposes, this category includes another 1.706 different patterns, although 1,526 of these patterns occur only up to 10 times (955 activity patterns appear only once in the data set). Within the more aggregated approach, 280 pattern types are summarized in the category ‘Other’, with 213 of them occurring only up to 10 times (100 activity pattern types occur only once). Another interesting fact of this analysis is that

55.6% of all pattern types are simple single tours with only one activity (‘HXH, ‘HYH’). It can also be recorded that 79% of all pattern types (excluding type ‘Other’) are based on simple single tours or simple multiple tours with only one activity per tour (e.g., ‘HYHYH’ or ‘HXHYH’). This is an important result as it demonstrates that for a great share of patterns the interdependency between activity locations and modes is not an issue due to their simplicity (only one activity and two related trips). Nevertheless, in case of simple multiple tours, time dependency is of relevance; for instance, time spent on the first tour affects the decisions taken on a second tour. Preliminarily we can conclude that travel behavior reproduced in the EOD is very multifarious; at the same time, only some simple patterns represent the great majority of the observed behavior. These dominating patterns are often characterized by only one or two activity stops per tour. On the other side, a large number of different and disperse patterns represent a minor share in the survey (often single entries). Based on these observations, in the next step the activity patterns are:

- Categorized according to the number and type of constituting tours.

The first step of the analysis already indicated further criteria available for their categorization. For instance, the pie charts differentiate between patterns with ‘Work’ or ‘Education’ activity and without it. They indicate that single tours are very often realized without being combined with additional activities. Taking the three major shares in Figure 4.11 represented by ‘HXH’, ‘HYH’ and ‘HXHYH’, we see that although each share represents an individual pattern type, ‘HXHYH’ can be further decomposed into two tours of ‘HXH’ and ‘HYH’. This rather trivial observation is of importance because a methodology to calculate the mode destination probability for pattern type ‘HXH’ will not differ from the one applied to the first tour of pattern type ‘HXHYH’. Additionally, one may assume that the summed activity times of the latter pattern type are longer than the time spent on the ‘HXH’ type, since there is a second tour included. In addition, we assume that the time already spent on the first activity ‘X’ would have an influence on the time spent on activity ‘Y’ in the second tour. These examples demonstrate that both the pattern and the tour perspectives are of relevance, as dependencies exist not only between single activities – and the associated mode and location decisions – but also between entire tours.

With the following Table 4.3 we provide a typology of all EOD activity patterns. The pattern types introduced in Figure 4.11 (right) are typified according to the following criteria: the number of constituting tours; whether a primary activity of ‘Work’ or ‘Education’ builds part of the pattern or not; and the number of additional activity stops made beside the primary ones.

Table 4.3: Typology of activity patterns

APT category	Example	Description	Work or Education	Number of Tours	Additional stops on Tour	Frequency	Percent	Cumulative Percent
1	HXH	Simple single tour	yes	1	0	13,245	31.6	31.6
2	HYH		no			10,055	24.0	55.6
3	HXHYH	Simple multiple tour	yes	2	0	3,120	7.4	63.0
4	HYHYH		no			4,735	11.3	74.3
5	HXHYHYH		yes	3		705	1.7	76.0
6	HYHYHYH		no			1,804	4.3	80.3
7	HYHXHYHXH		yes	4		129	0.3	80.6
8	HYHYHYHYH		no			497	1.2	81.8
9	HXYH	Complex single tour (1)	yes	1	1	1,270	3.0	84.8
10	HYYH		no			1,328	3.2	88.0
11	HXYHYH	Complex multiple tour (1)	yes	2	1	488	1.2	89.2
12	HYHYYH		no			764	1.8	91.0
13	HXYHYYH		yes		2	26	0.1	91.04
14	HYYHYYH		no			67	0.2	91.2
15	HXHYYHYH	Complex multiple tour (2)	yes	3	1	42	0.1	91.3
	HYHYXHYH					51	0.1	91.4
	HYHXHYYH					40	0.1	91.5
16	HYHYYHYH		no			153	0.4	91.9
	HYHYYHYH					108	0.3	92.1
	HYHYHYYH					107	0.3	92.4
17	HXYHYHYYH		yes		2	8	0.02	92.41
	HYHXYHYYH					4	0.01	92.42
	HYYHXYHYH					7	0.02	92.44
18	HYYHYHYYH		no			15	0.04	92.48
	HYHYYHYYH					11	0.03	92.50
	HYYHYYHYH					16	0.04	92.54
19	HYXYH	Complex single tour (2)	yes	1	2	980	2.3	94.9
n.c.	HYXYYH				3	244	0.6	95.5
	HYYXYYH				4	120	0.3	95.7
	HYXYYYYH				5	33	0.08	95.8
	HYYXYYYYH				6	16	0.04	95.9
	HYYYH		no		2	519	1.2	97.1
	HYYYYH				3	206	0.5	97.6
	HYYYYYH				4	59	0.1	97.7
	HYYYYYYH				5	21	0.05	97.78
	HYYYYYYYH				6	13	0.03	97.81
n.c.	HXHYYYH	Complex multiple tour (3)	yes	2 or 3	>= 2	446	1.1	98.9
	HYHYYHYYH		no			470	1.1	100.0
	Total					**41,922**		

Annotation: The column ‚APT category' reflects those categories recognized later in the estimation of activity pattern probabilities (see section 6.1). The abbreviation 'n.c.' means 'not considered'. All patterns with more than three additional stops and those described as 'Complex multiple tour (3)' are excluded.

Each activity pattern is composed by one or more tours. As stated above, further criteria for the building of the typology were considered that led to the definition of seven pattern types. These are characterized as follows:

1. Simple single tour: pattern is based on one tour, one activity and two trips; no additional stops (activities) are realized.
2. Simple multiple tour: pattern is based on two, three or four tours; each tour contains one activity and two trips; no additional stops (activities) are realized.

3. Complex single tour (1): pattern is based on one single tour, two activities and three trips; one additional stop (activity) beside the primary activity is realized.
4. Complex multiple tour (1): pattern is based on two tours; one or two additional stops (activities) per tour are realized.
5. Complex multiple tour (2): pattern is based on three tours; one or two additional stops (activities) per tour are realized.
6. Complex single tour (2): pattern is based on one tour; two to six stops (activities) are realized.
7. Complex multiple tour (3): pattern is based on two or three tours; two ore more additional stops (activities) per tour are realized.

The maximum number of consecutive tours is four since the analysis of the data set was limited to activity pattern with maximum seven out-of-home activities (see 4.3.1). The decomposition of activity pattern into tours leads to an average of 1.44 tours per day and per person (60,191 tours constitute 41,922 activity patterns). The following Figure 4.12 illustrates the predominance of some of the grouped pattern types and shows shares of each of the seven grouped pattern types as indicated in the table above.

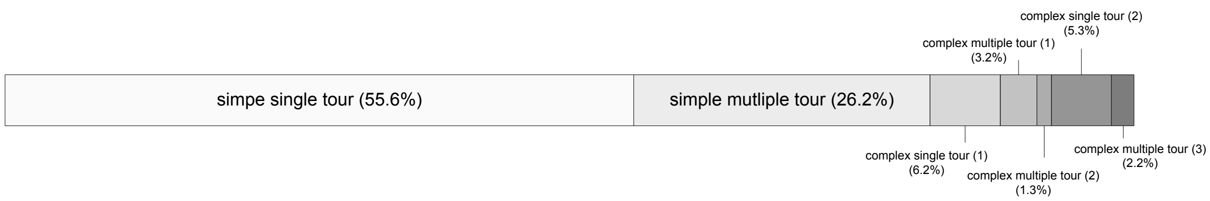


Figure 4.12: Shares of grouped activity pattern types

On the whole, the typology supports the idea to focus on less complex pattern and tours. Nearly 56% of all patterns are simple single tours with only one out-of-home activity. These tours do not require estimation of dependencies in timing or in activity location choice. Another 26% are simple multiple tours that also contain no more than one activity but with several tours in a row. Including the complex single tours and the complex multiple tours (1), already 91% of all patterns are characterized by one or more tour but with no more than two additional stops.

Following the observation that the great majority of activity patterns consist of no more than two consecutive activities in between home stays, it seems reasonable to rely the empirical application on pattern type ‘HXYH’. This tour/pattern includes all patterns that are composed by macro activities of work or education (‘X’) and where a secondary activity is ‘Shopping’, ‘Leisure’ or ‘Other’ (‘Y’). ‘HXYH’ is responsible for about 3% of all patterns reported in the EOD (see Figure 4.11). But more importantly, this pattern is suitable for the analysis since the estimation of the two scales activities ‘X’ and ‘Y’ in time and space requires explicit treatment of the relationship between macro and micro scales decisions. The pattern type ‘HXYH’ and the respective dependencies serve as a sort of a ‘prototype’. We assume that provided the decisions are traceable in a probabilistic fashion for this tour, many other relatively similar tours (e.g., ‘HYXH’ or ‘HYYH’ types) and patterns can be calculated with an adapted or extended methodology. Primarily due to limitations of the sample size, the approach picks up this pattern in an aggregated manner. Nonetheless, the methodology is not dependent on this specific pattern type, meaning that, for instance, a more disaggregated pattern, such as ‘HWLH’, can be estimated as long as the empirical data is available.

4.4 Summary

This chapter served as a transition between the previous chapter, where theory and methodology were introduced, and the following chapter that focuses on the statistical data analysis. To this end, the disaggregation of the urban core into 618 Traffic Analysis Zones and more than 50,000 residential blocks was introduced. Santiago’s land-use and transport models were described as they provide important data and information. Namely, the land-use of service and commerce activities as well as the number of user groups per TAZ provided by land-use model MUSSA are used. Transport model ESTRAUS delivers the commuting matrices by user group and mode and the level-of-service matrices by mode. This information is taken as granted when it comes to the estimation of spatial path flows.

The next section exemplified information available by the EOD data set. Data mining of the raw data set was necessary, and reference was given to the detailed preparation steps which are documented through a set of syntaxes described in Annex-Table 6. To assure that the reduced data set still was representative for Santiago’s population, we compared raw and reduced data sets and came to the conclusion that no unsystematic elimination of cases occurred. The subsequent analysis concentrated on showing different approaches discussed in literature of how to determine primary importance among activities. Applying the two-level hierarchy approach we found out that on a regular workday shopping activities showed the highest relevance after the top-priority work and education activities, followed by a relatively even distribution between other discretionary activities. In addition, the two-level hierarchy provided almost the same result as if we assigned activity priority only dependent on duration. We found that to be in line with the theoretical assumption that the resources invested (here time, i.e. activity duration) are related to the activities scale.

We then created a typology of activity patterns and tours. In its final version, this typology is built on aggregated travel purposes differentiating only between work and education, the ‘X’ activity, and any other activities, denominated as ‘Y’. This simplification and the introduction of a set of criteria allowed identifying seven major types of activity pattern. It was observed that the variety of activity patterns is tremendously high in the data set, but with regard to quantitative representativeness very few patterns dominate the picture. According to the typology, the chosen example of pattern type ‘HXYH’ allowed outlining its relevance among other patterns. This relevance does not refer to its actual share in the total of observations but to its character of representing the conditionality of choices, including both macro and micro scale decisions. The following chapter will concentrate on the issue of time-space constraints. We will describe the constraints for the case of tour-type ‘HXYH’, conducting the empirical analysis using the data set prepared in this chapter.

5 Empirical analysis of time-space constraints

The main objective of this chapter is to identify time-space constraints and to estimate thresholds that allow the reduction of potential decision options. The issue of a constrained environment and its impact on the choices is dealt with in the model in different aspects. The conditionality between different scales at which decisions are taken was subject of the chapter on the methodology. Macro decisions influence or, in other words, constrain subsequent decisions, for instance, those regarding mode and location choice. Another constraint is time. Time as a limited personal and daily resource naturally constrains the spatial area where travel destinations are located. Third, there is a cognitive limit of individual's ability to process any theoretically available option. The cognitive factor indicates that not all of a huge number of options available play a role when it comes to the travel related choice.

All these aspects are subject of the empirical analysis that follows. It appears interesting to investigate whether the theoretical assumptions about hierarchical decision-making in a constrained environment are reflected in the data. The analysis deals with the following issues (in the following order):

- **Dependency of mode choices.** The related question is: what are the mode combinations that should be considered when modeling pattern type 'HXYH'? The identification of mode combinations is subject of the first section. It determines number and type of combinations considered in the model.
- **Total daily travel time.** It was argued that time as a limited resource constrains behavior, i.e. travel-related decisions. The subject of the second section is the identification of daily travel time thresholds, which reflect the maximum time an individual (or user group) is likely willing to spend on travel per day. These thresholds are applied as parameter in the model and reflect the time constraint.
- **Daily activity times.** This section deals with the identification of time regimes, i.e. with starting times and duration of activities 'X' and 'Y'. Grouped time periods on an hourly basis are one of the results of the analysis. In addition, we check whether socio-economic or geographical variables do have an influence on the identified time regimes.
- **Detour factors.** The following section deals with the calculation of detour factors that allow defining a geographical area where secondary activities are most likely to occur. This section is the most extensive one, as we first introduce the concept of detours in general terms and after that demonstrate how they can be calculated using the EOD data. These descriptions also include statistical analysis via regression models. The obtained factors are used for the estimation of search spaces; thus, they represent the space constraint.
- **Local attraction.** The final section starts with a brief introduction of the land-use data considered, introduces the distinction of home- and work-related tours and illustrates the travel times considered by impedance matrices.

At first sight the upcoming sections appear relatively independent, and their interrelation might be hard to grasp. However, each section represents an important component of the model, for instance, empirical thresholds or suggestions for grouping behavioral heterogeneity. The interrelationship of all issues treated in this chapter as well as their concrete function within the approach becomes even more tangible in the subsequent chapters that present the results.

5.1 Dependency of mode choices

The calculation of trips throughout a day requires the dependencies between decisions for transport modes to be considered. Staying with the example of pattern type 'HXYH', three trips are realized without a home stay in between. Considering four transport modes, this results in a possible number of mode combinations of 81 (3^4). However, it does not seem reasonable to consider the complete variety of theoretically possible mode combinations within the application. Certain mode combinations might be preferred as options and thus show higher frequencies of occurrence. For instance, if the first trip is done by car (as a driver), only in very few cases the adjacent trips will be done by other modes because the necessity to bring the car back determines the behavior and the respective associated decisions.

The decision to select only a part of the overall variety of possible mode combinations is taken considering the following criteria: first, for each transport mode (public transport, car driver, car passenger, walking and bicycle) at least one mode combination should be included. Second, for each of these principal modes we include one 'pure' mode combination referring to the use of the principal mode throughout the pattern. Third, the total number of mode combinations incorporated in the approach should represent a great majority of the overall observed variations to assure representativeness.

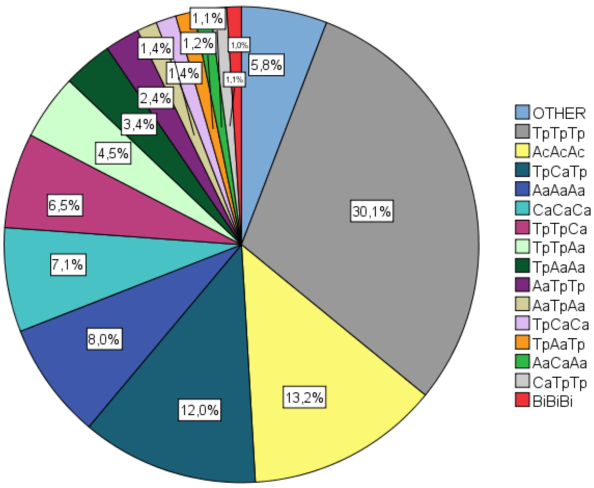


Figure 5.1: Share of mode combinations, pattern type 'HXYH'

Annotation: The abbreviations used in the legend stand for the principal modes: Tp (Public Transport), Ac (Car driver), Aa (Car passenger), Ca (Walking) and Bi (Bicycle). By far dominant and reflecting a 43% of all variations are the combinations of TpTpTp and AcAcAc. In total, 86 mode combinations were reported in the EOD given that the initial analysis was realized based on a disaggregated mode variable, which included also 'shared taxi', 'taxi' and 'other' modes. These modes with comparably small shares were assigned to the principal modes ('shared taxi' and 'other' to Tp, 'taxi' to Aa), reducing the overall variety to 36 mode combinations. 21 of the 36 combinations were reported less than 1% of the times mentioned and were grouped into the category OTHER. / (N=805)

The 15 mode combinations of Figure 5.1 represent about 94% of all reported mode combinations for pattern type 'HXYH'. The selection introduced here fulfills the criteria mentioned above: every principal mode is included, a respective 'pure' mode combination could be considered for each mode and the great majority of combinations (94%) is reproduced. Theoretically the approach is able to consider even more mode combinations. However, due to the limited number of observations for other mode combinations we will not handle all the behavioral choices observed. For instance, the mode combination of type AaCaCa, responsible for less than 1% of all observations and belonging to the category OTHER in Figure 5.1 is not considered, thus no spatial paths probabilities will be calculated for this specific mode combination.

At this point, the connection between the empirical analysis of the EOD and the information used from Santiago's transport model ESTRAUS become evident. Concretely, every OD flow between home and work locations provided by ESTRAUS will be later disaggregated by the 15 mode combinations derived from the EOD analysis. In other words: any trip from home to work that is currently interpreted as a single and independent event (trip-based) will be disaggregated regarding the identified probabilities for adjacent modes (tour-based). For example, if a trip to the primary activity is done by public transport, several possibilities open up that second and third trips are conducted by other modes. For each principal mode and according to the number of subsequent mode combinations, tour-based modal split values for each combination are defined. The Figure 5.2 shows the distribution of principal modes into mode combinations and its related disaggregated modal split values.

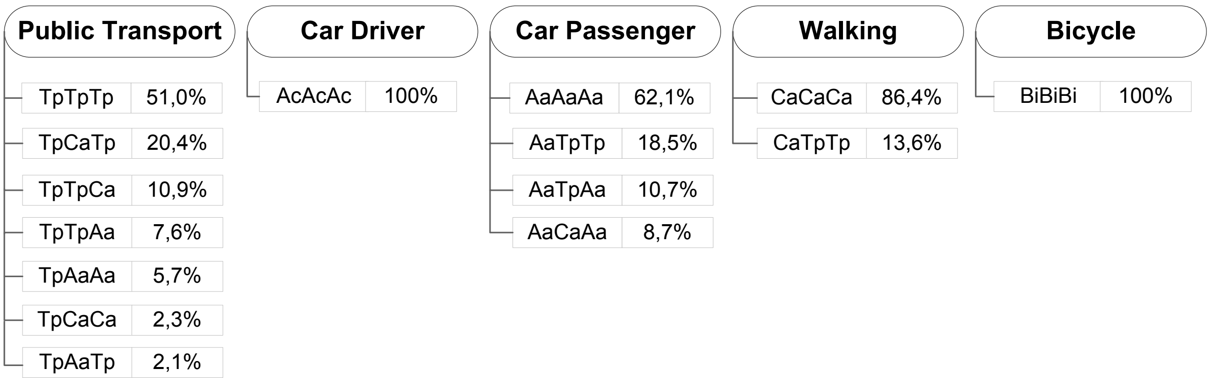


Figure 5.2: Tour-based modal split values / (N=758)

Figure 5.2 clarifies two major issues: a) the dominance of 'pure' mode combinations where, for example, in 86.4% of all combinations walking mode, if started, continues throughout the pattern and b) the heterogeneity of mode choices is most significant when the trip between home and work is done by public transport. There, 51% of all combinations are realized using only public transport modes, whereas other six combinations indicate options where modes are mixed. Notice that additional criteria of socio-economic character or the availability of cars in household user groups – aspects that have an influence on mode choices – have not been included so far. In a subsequent analysis the tour-based modal split values were estimated considering the socio-economic user groups, given the condition set by the 15 mode combinations.

The corresponding analysis considering household user groups indicates the discrepancies due to income and availability of cars. The following Figure 5.3 displays the observed shares of the 15 mode combinations by 13 household user groups.

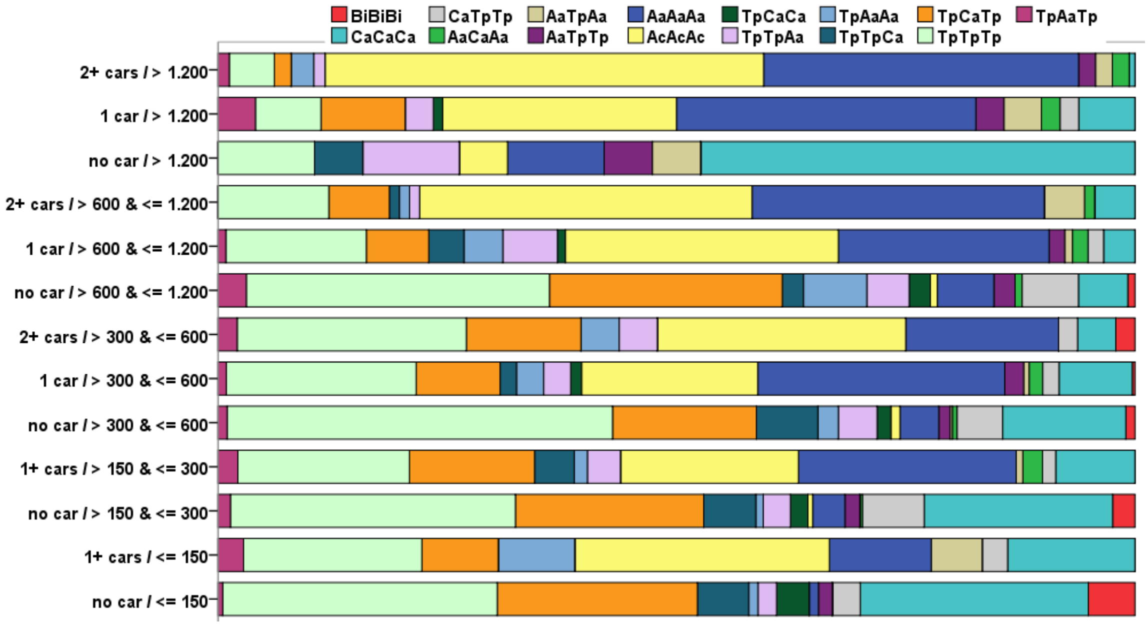


Figure 5.3: Mode choice combinations by household user groups

Annotation: To be able to realize the analysis on more cases, the activity pattern types 'HXYH', 'HYXH' and 'HYYH' were considered. This is defensible against the assumption that in each of these complex single tours similar decisions had to be made, for the primary and secondary activities and the conditioned choices on modes. Notice that also for 'HYYH' a primary activity is defined according to the two-level hierarchy introduced before assigning the longest activity in duration a primary status. / (N=2,226)

Some specific distinctions can be deduced from Figure 5.3. For example, the mode combination TpTpTp is the most frequent type among all user groups with no car available (between 29.9% and 42%), except for the user group with the highest income. An exceptionally high share of CaCaCa type is observed (47.4%), which might be a partially biased result due to a small amount of observations in this group (19 observations). In user groups where one or more cars are available, the mode combinations AcAcAc and AaAaAa dominate (between 23.7% and 47.9%). The use of bicycles indicates a significant share in particular in the lowest income group, but it should be noticed that in 2001, at the time of the survey, the use of bicycles did not play an important role in Santiago. Other relevant combinations are those where a trip to work by public transport is combined with other modes (TpAaAa, TpCaTp, TpTpAa and TpTpCa). In these cases non-motorized households dominate, but the use of public transport is more evenly distributed among all user groups. Notice that differentiated shares of mode combinations by household user groups enter the calculation approach (regarding the detailed mode choice combination probabilities, refer to Annex-Table 4).

5.2 Total daily travel time

The time constraint considered in the model is the total daily travel time. It was argued in the introductory chapters that time for activities and travel is a limited resource, and individuals make decisions according to restrictions and constraints. Respectively and based on the previous analysis of the mode combinations of pattern type ‘HXYH’, the question is to what extent total daily travel time depends on and varies according to the mode combinations. One may expect differences between motorized and non-motorized modes as well as between public and private transport.

The criterion ‘maximum daily travel time’ is used to reduce the number of feasible secondary activities locations to be reached given the travel time available. This criterion is implemented in the model in such way, that no spatial paths will be estimated that show a daily travel time beyond the identified thresholds by mode combination. One has to be aware of the fact, that due to data diversity uncommon cases, or outliers, are responsible for very high travel times. The existence and effect of outliers reflect a general concern of this work, where the focus is on modeling of the principal behavioral pattern, rather then of seldom or randomly occurring events. The following Table 5.1 summarizes the statistics for the daily travel times for pattern type ‘HXYH’ by the 15 mode combinations introduced in the previous section.

Table 5.1: Daily travel time by mode combination

		Daily Travel Time (pattern type 'HXYH')								
Nr	Type	Valid N	Minimum	Maximum	Mean	Median	Percentile 80	Percentile 90	Percentile 95	Standard Deviation
1	TpTpTp	242	37	300	141	135	185	215	233	52
2	AcAcAc	106	18	205	91	90	120	140	150	38
3	TpCaTp	97	35	250	110	105	148	170	191	47
4	AaAaAa	64	17	285	68	60	95	105	125	43
5	CaCaCa	57	12	145	38	30	50	68	105	27
6	TpTpCa	52	18	230	115	112	150	180	220	50
7	TpTpAa	36	34	280	123	118	165	180	220	51
8	TpAaAa	27	36	197	88	85	110	150	155	39
9	AaTpTp	19**	50	195	103	100	128	140	195	35
10	AaTpAa	11**	35	130	84	90	105	129*	130	33
11	TpCaCa	11**	35	104	64	60	68	95*	104	20
12	TpAaTp	10**	48	221	128	122	141	190*	221	50
13	AaCaAa	9**	30	121	67	70	86	111*	121	33
14	CaTpTp	9**	37	170	97	100	120	143*	170	43
15	BiBiBi	8**	18	160	70	68	78	104*	160	43

Annotation: ordered by column ‘Valid N’, combinations with less than 20 observations are flagged with ** / (N=758)

Generally, daily travel times have the highest values (mean, median, percentiles) when two or more public transport trips are realized (Nr. 1, 3, 6, 7, 9, 12 and 14). The lowest values are detected for the two-mode combinations containing two or more walking legs (Nr. 5, 11). In between is the pattern with two or more car-based trips (Nr. 2, 4, 8, 10 and 13). Standard deviation is at highest among the public transport based, and at lowest among walking-based combinations. This reflects the greater range of daily travel times among motorized trips, whereas walking-based daily travel times vary less. These general tendencies persist even in case of mode combinations with a low number of observations. The comparison of percentiles 80, 90 and 95 allows identifying the effect of outliers. In the case of TpTpCa (6), TpAaTp (12) and AaTpTp (9) this is most evident with a difference of up to 80 minutes between the 80 and 95 percentiles. To reduce the influence of outliers, in a first stage, the percentile 90 was chosen for all mode combinations to determine the maximum daily travel time. In addition, the mode combinations with less than 20 cases were analyzed looking in detail at the order of values of all cases (because of the low number of observations). Except the mode combination AaTpTp (19 cases), all other combinations with less than 20 observations indicated some sort of 'leap' looking at the highest values reported. For instance, the percentile 90 of mode combination TpCaCa is 95 minutes and the percentile 80 is 68 minutes with no other reported values in between. In these cases, flagged with *, instead of the percentile 90, the values of percentile 80 are chosen and considered in the model.

The decision to consider 15 mode combinations made in the previous section resulted in ignoring of 6% of heterogeneity. Analogously, the implementation of a maximum daily travel time excludes very long travel times (considering the thresholds defined by the percentiles 90 and 80, respectively). For instance, no spatial paths will be calculated that exceed an overall daily travel time of 215 minutes in the case of TpTpTp mode combination or 68 minutes in the case of CaCaCa type.

5.3 Daily activity times

Time spent on activities conducted during the day is integrated into transport models in quite different ways: the classical four-step models generally do not consider time along with the estimation of demand; hence, the processes of generation, distribution and mode partition are free of temporal aspects. The differentiation of the demand by time period (e.g., peak hours) is often implemented afterwards by the division of the daily demand into demand shares according to empirical observations. Applications belonging to the family of activity-based approaches often consider time during demand estimation, interpreting temporal aspects as decisive in the choice process for activities and trips. However, any application needs to specify how time aspects are integrated, for instance, if the time period for demand estimation refers to a single day or a week and if starting times or duration of activities and trips are defined in minutes, hours or other time units that appear useful.

5.3.1 Grouping of activity starting times and duration

In case of dependencies between a macro and a micro scale dealt with in this analysis, these requirements apply to the starting times and duration of the out-of-home activities 'X' and 'Y'. Respectively, the questions are as follows. When does the primary activity ('X') start? How long will it take? Given a starting and duration time for 'X', what can be expected regarding the duration of the secondary activity ('Y')? In consequence, the reproduction of

'HXYH' requires empirical specification of these temporal dependencies. Dependencies are defined via the priorities between macro and micro activities. This means, the temporal decision of when to conduct the primary activity, determines characteristics, particularly, the duration, of the secondary activity. Hence, the empirical analysis follows the theoretical concept of macro and micro scales, where the starting time and duration at the micro scale – the 'Y' activity – depend on the decisions taken at the macro scale – the 'X' activity. But, dividing the day into 24 hours, theoretically 24^3 (13,824) temporal combinations are possible for a plan of type 'HXYH'. Temporal combination is defined as a probability that the secondary activity takes, say 3 hours, given the duration of the primary activity is 9 hours, and so on. To consider all combinations is neither feasible nor necessary from a behavioral point of view as it can be expected that some combinations appear only seldom. For instance, 1-hour duration of the primary activity, followed by 13-hour duration of the secondary activity will rarely be observed in the data. Respectively, the empirical analysis has to concentrate not only on the temporal dependencies between the macro and the micro scale but also on the identification of the relevant time periods at each scale considered. The objective is then to identify the most relevant periods that allow reducing the total number of combinations. Similar to the sections before, the main objective is to reduce the choice combinations still assuring that the major temporal combinations remain considered. Against this background, the following conditions apply for the analysis:

- The temporal unit for the analysis is an hour, both for primary and secondary activities.
- The starting time of the primary activity depends on its duration.
- The duration of the secondary activity depends on the duration of the primary activity.
- It is tested whether the analysis should be differentiated by a socio-economic or geographical influence variable.

Figure 5.4 and Figure 5.5 show to what extent starting time and duration regimes allow for the verification of the assumption, that time periods can be identified. The upper left figure shows the distribution of starting times of the primary activity 'Work'. The individual starting times of each analyzed activity pattern were grouped into an hour-based value, for instance, if activity started at 8.25h, it belongs to value '9' on the axis of starting hours (>= 08.01h & <= 09.00h). There is a sharp increase in the share of activities starting between 07.01h and 09.00h in the morning (60%). Even though the concentration of starting times can be confirmed, the upper left figure demonstrates small shares for nearly all hours during the day. In the upper right figure – based on the same sample of activity patterns – all patterns with a share of less than 1% were assigned to the margin hour categories. For instance, all patterns that started before 05.01h in the morning (14 patterns, in sum 0.2%) were assigned to the starting hour '6' (170 patterns, in sum 1.7%, those starting after 17.00h were assigned to starting hour '16'). The same criterion (< 1% share) was applied for the distribution of primary activity duration, shown in the figures in the middle left and right. There, 166 patterns with less then 121 minutes of activity duration (in sum, 1.7%) got assigned duration hour '3', meaning a duration of 121 to 180 minutes (2 to 3 hours), 74 patterns with more than 15 hours duration (in sum, 0.8%) got assigned duration hour '15', meaning duration of 841 to 900 minutes (14 to 15 hours). In case of duration of a secondary activity (bottom figures left and right), the method led to 5 remaining duration hours, grouping 21 patterns (in sum 2.9%) with a duration of more than 5 hours to the duration hour ‘5’.

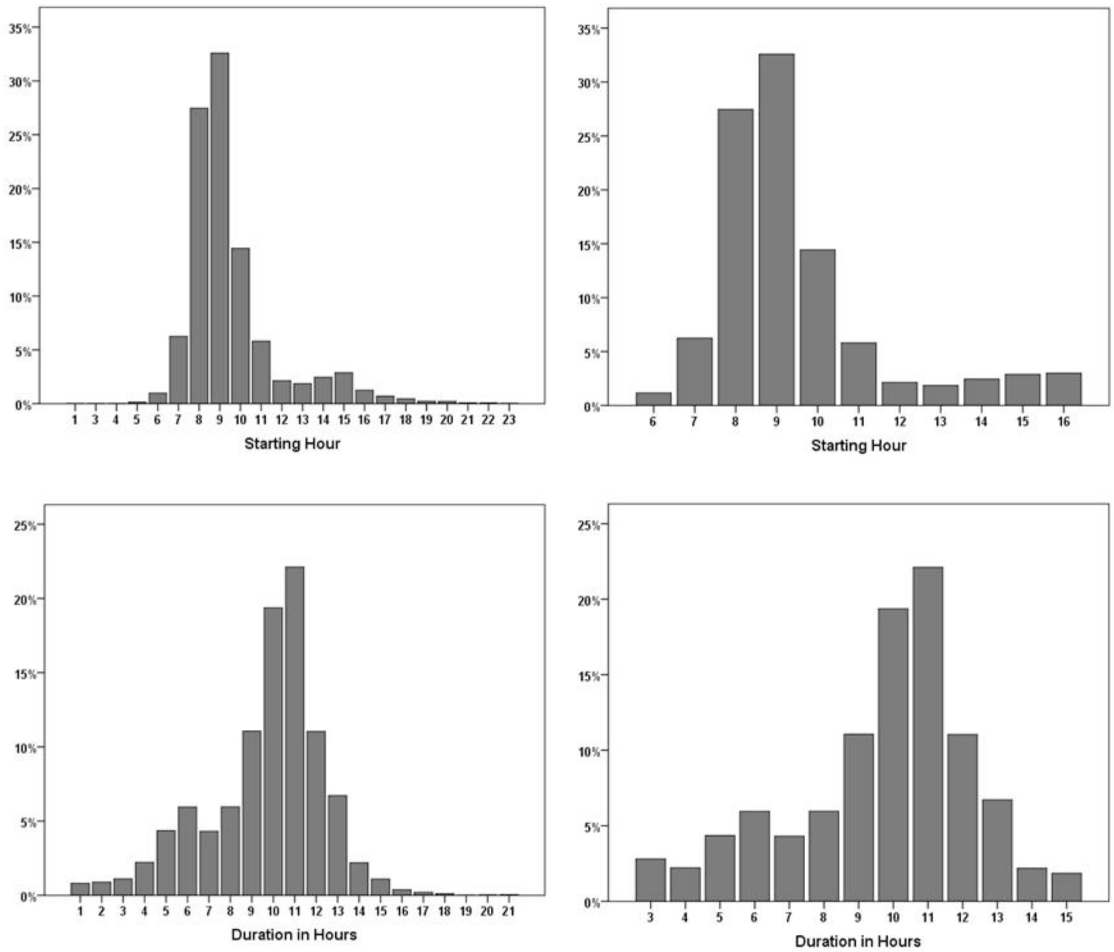


Figure 5.4: Starting time and duration for the primary activity.

Annotation: For both the analysis of starting time and duration of the primary activity, all activity patterns with one tour and a primary activity 'Work' were selected. This is legitimate according to the assumption that the choices associated to the primary activity (macro scale) are independent from the choices associated to the secondary activity (micro scale) within one tour. Left: disaggregated; Right: aggregated. / (N=9,753)

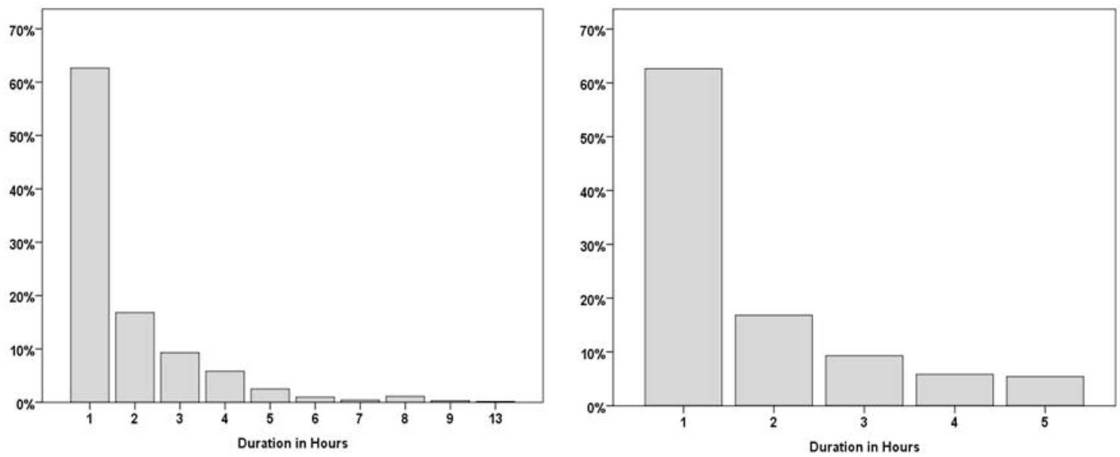


Figure 5.5: Duration of the secondary activity.

Annotation: For the analysis of the duration times of the secondary activity, all activity pattern types 'HXYH' and 'HYXH' were selected if 'X' was reported as work activity. Left: disaggregated; Right: aggregated. / (N=720)

5.3.2 Influence factors on time regimes

In addition to the previous analysis, the distribution of starting and duration times – now already based on the aggregated time periods – is conducted considering a socio-economic and a geographical variable. It is assumed that geography, the area where people live in the city, may have an influence on their temporal activity regimes. This might be due to distinct transport supply qualities or to various employment patterns, whether people start work earlier or later. Similarly, a socio-economic variable can play a role: for instance, households with on average higher income might have jobs with different work time regimes than households with lower income and lower level of employment. Respectively, the analysis is based again on the user group variable (socio-economic variable) and Santiago's city sectors (geographical) (see Figure 5.6). If a correlation between theses variables and the time regimes were found, this would support the idea of integrating them in the calculation procedure. The Figure 5.6 below presents visually the distribution of starting time regimes by city sectors and user groups (Figures for the relationship of primary activity duration times by city sector and user groups were moved to Annex-Figure 1).

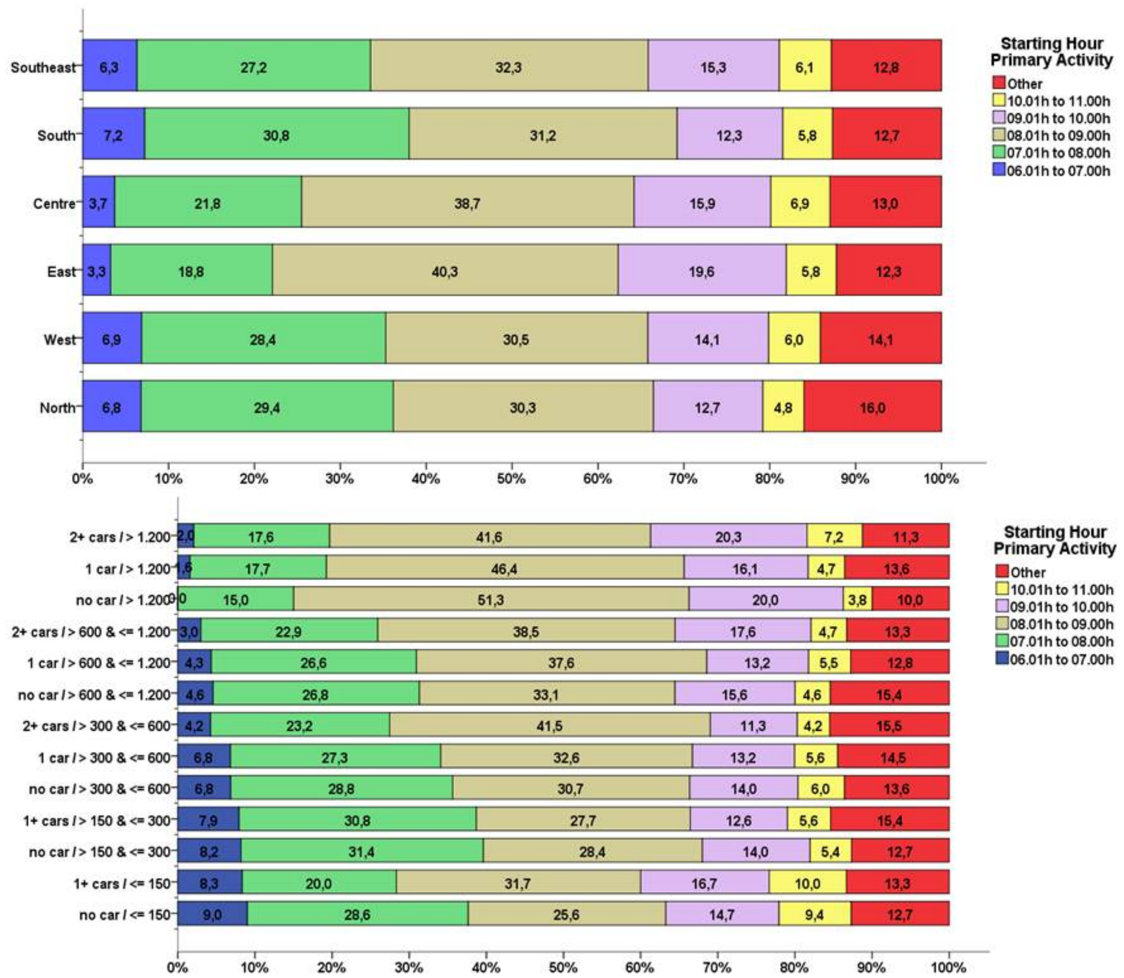


Figure 5.6: Starting hour of primary activity by city sector and user groups

Annotation: Pictured are the percentages of observations. The category 'Other' comprises all shares of starting hours with less than 3%. These were grouped due to ensure visibility among the dominating starting hours in the morning. / (N= 9,753)

Both figures visualize similar interesting findings. Especially for the early starting hours between 6h and 9h the differences become obvious. Regarding city sectors, the center and the eastern sectors indicate a distinct pattern with significantly lower shares of early starting hours, showing higher shares of starting times between 9h and 10h. This phenomenon is slightly more pronounced in the eastern city sector. The same phenomenon is observed in case of the user groups but is related to income and car ownership. With increasing household income there is a tendency for the starting times for primary activities to shift towards later hours. Within the same income group this tendency seems to be reinforced by car ownership, where the possession of a vehicle increases the likeliness of starting the primary activity later. These findings become comprehensible given the spatial structure of income distribution in Santiago. The city is highly segregated with the affluent neighbourhoods living in the central and particularly in the eastern parts of the city. To some extent, it can be assumed that both variables 'tell the same story', indicating different time regimes for those 'better off'. Apparently, differences in working conditions or job types of residents of the southern and western parts of Santiago make them start work earlier. As these differences in the time regimes seem significant, it appears rational to consider the variables when it comes to integrating them in the model. But first we should verify to what extent they correlate and whether it is sufficient to consider one of the variables. Thus, possible combinations were tested for statistical correlation. The following Table 5.2 summarizes the results, with the left column indicating the combinations tested.

Table 5.2: Correlation between time regimes, user groups and city sectors

		Chi-Square	Contingency Coefficient	Significance	% cells with < 5 counts	Valid N
1	Starting hour primary activity by city sectors	200.9	0.142	0.000	1.5	9,753
2	Starting hour primary activity by user groups	342.2	0.184	0.000	17.5	9,753
3	Starting hour primary activity by income	250.6	0.158	0.000	0.0	9,753
4	Starting hour primary activity by car ownership	66.8	0.082	0.000	0.0	9,753
5	Duration hours primary activity by city sectors	136.6	0.118	0.000	0.0	9,753
6	Duration hours primary activity by user groups	301.5	0.173	0.000	12.4	9,753
7	Duration hours primary activity by income	158.6	0.126	0.000	0.0	9,753
8	Duration hours primary activity by car ownership	49.7	0.071	0.000	0.0	9,753
9	Duration hours secondary activity by city sectors	17.5	0.154	0.618	13.3	720
10	Duration hours secondary activity by income	23.3	0.177	0.106	12.0	720
11	Duration hours secondary activity by car ownership	11.0	0.123	0.026	0.0	720
12	City sectors by user groups	2,690.2	0.465	0.000	2.6	9,753

Chi-squares are highest for the combinations considering the user groups. To blind out the chi-squares dependency on the sample size, the contingency coefficient is estimated. The highest coefficients at a significant level show combinations where user groups were considered (0.184 and 0.173). Coefficients are in both cases (starting and duration hours) higher – indicating a stronger relationship between user groups and time regimes – than for the same analysis considering the city sectors. When analyzing correlations between the duration of secondary activity and user group many cells remained with less than five cases. In consequence, the user group variable was aggregated; first grouping the income categories and second grouping the car ownership (see Table 4.1). As a result, significant correlation was found for the car ownership variable while the income variable was significant only at a 90% confidence interval. The same analysis with reference to starting time and duration of the primary activity revealed a stronger correlation with the income than with the car ownership variable (starting hour: 0.158 against 0.082; duration hours: 0.126 against 0.071). The test for correlation between the city sectors and user groups confirmed the assumption expressed above that both variables tend to explain the same phenomenon. Contingency coefficient is significant and high with 0.456. In consequence, the following conclusions are drawn:

- The differentiation by user groups has a significant impact on the starting time and duration of the primary activity. But this effect is primarily due to income disparities, whereas car ownership plays a less important role.
- No correlation was found between duration of the secondary activity and the city sectors. A significant effect could be stated for car ownership.
- The assumption that both the city sectors and user groups explain partly the same phenomenon is confirmed.

Notice that in this analysis only one socio-economic and one geographical variable are considered. In addition, user groups are composed according to two criteria (car ownership and income) that both tend to represent a households' economic situation rather than constraints related, for instance, to the family structure. The consideration of additional variables may lead to the identification of further variables that have an effect on the time regimes, for instance, age, gender, or children in the household. We focus here on purpose on user groups and city sectors since these variables are part of the model structure; thus, their effects can be considered in the calculations. A more detailed analysis of other variables is recommended once they are represented in the model.

In consequence to these findings, the assignment of time regimes later in the model follows the empirical distributions of primary activities duration considering user groups (regarding aspects of implementation, see section 3.2). The distributions of starting times of the primary activity and duration of the secondary activity are dependent on the duration of the primary activity, thus on the differentiation by user groups. As it was stated above, city sector as a variable is rejected in favor of user groups.

Notice that the decision to condense the time periods as shown in Figure 5.4 and Figure 5.5 to those most important reduces the complexity but at the cost of losing some behavioral variation observed in the data. This leads to a loss of 1.9% of variation in case of starting times of the primary activity and to a loss of 2.9% of variation in case of duration times. The reduction of duration of the secondary activity to a maximum of 5 hours reflects a loss of 2.9% of the behavioral heterogeneity. The definition of time regimes is picked up in the subsequent analysis, which deals with detour factors to define potential search spaces for secondary activity locations. There, inter alia, the assumption is tested whether search spaces depend on the time spent on the primary activity.

5.4 Detour factors

So far, the only criterion that limits the number of potential location options is the maximum travel time needed to realize a pattern of type 'HXYH'. A spatial constraint is introduced below, which delimits the spatial extension of the activity search space. For this purpose, we first introduce the conceptual idea of detour factors and then proceed to an empirical analysis focusing on providing a set of detour factors used for the estimation of search spaces.

5.4.1 Conceptual background

In literature, detour factors are denominated as a comparative value between a reported distance and some kind of reference distance, for instance, the actual travel distance between an origin and a destination and the respective straight-line distance (Witlox, 2007, p. 174; Zumkeller et al., 2005, p. 76-7). In general, factors can be generated for temporal and spatial detours (Dugge, 2006, p. 111). Witlox (2007) specifies the three most commonly used computed distance measures: a) the straight-line distance (SD), b) the shortest-distance path distance (SPD) and c) the shortest-time path distance (ST_1T_2D). The following Figure 5.7 illustrates these distance measures.

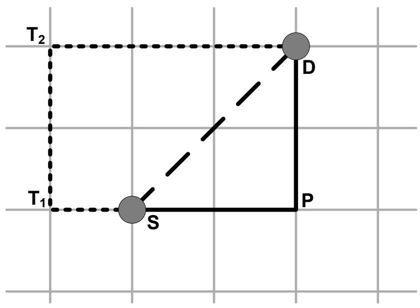


Figure 5.7: Distance measures

Annotation: Following Witlox, 2007, p. 174 / S = start point, D = end point

The measures SPD and ST_1T_2D require information about networks and time needed to move on the network, whereas SD requires only geographical locations of S and D. Any of these distance measures represent an approximation to what would be the travel distance in real life. In this sense SPD and ST_1T_2D are more realistic since they consider aspects of the transport system by incorporating network characteristic whereas SD ignores them. Nevertheless, an individual route choice seems to integrate many more criteria and is not always minimizing the respective costs (Witlox, ibid.).

Here, the concept of detour factors describes a detour (or a deviation) that an individual makes to conduct a secondary activity. Given the pattern type 'HXYH', it can be expected that activity 'Y' requires a detour from the direct way between home and work. According to the illustration below (Figure 5.8) this means that the distance sum of DP and PS (t_2+t_3) is expected to equal or to exceed SD (t_1).

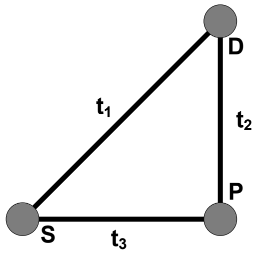


Figure 5.8: Illustration of the detour concept

Given the distance measures introduced, detour factors may theoretically be estimated by applying the measures SD, SPD or ST_1T_2D as suggested in Figure 5.8. But the EOD, which makes up data source for the estimation of detour factors, provides only location coordinates of origins and destinations, but not of specific routes chosen. Ideally, factors are estimated using *t* as the travel time on actual network routes chosen, because compared to topologically defined factors they are sensitive to congestion levels. Indeed, considering networks is expected to improve the measurements of the detour factors, especially in tours by public transport or car. There is a possibility to combine the information of origins and destinations of the EOD with transport networks in order to estimate SPD or ST_1T_2D. However, this option is not considered here since it appears difficult to estimate the respective workload that would not necessarily contribute to a conceptual improvement of the methodology itself.

In consequence we decide to apply the more simple straight-line distance measure (SD) for further calculations. For each 'HXYH' pattern in the EOD the SD for t_1, t_2 and t_3 is calculated. The following Figure 5.9 demonstrates this on an example of a Home-Work-Shopping-Home tour, with 'Home' located in the municipality of Renca, 'Work' located in Providencia and the shopping activity in Santiago. The underlying smaller geographical units reflect the TAZ.

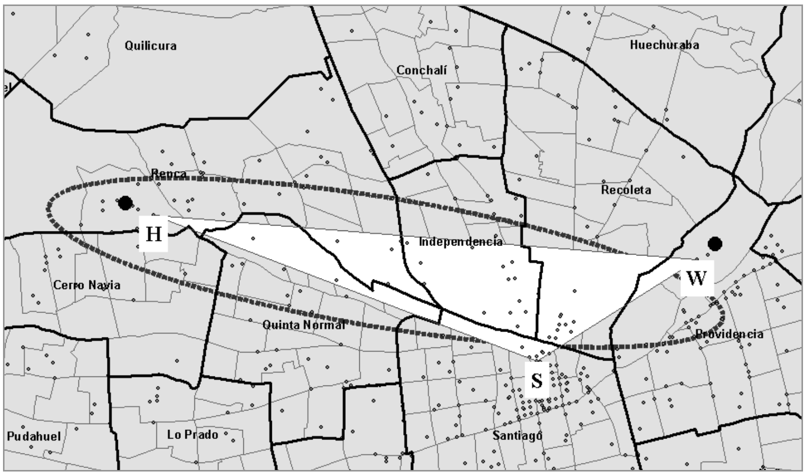


Figure 5.9: Examples: minimum convex polygon and standard deviation ellipse

Smaller points indicate further secondary activity locations of other 'HXYH' patterns reported in the EOD. The example also shows that due to the detour to the secondary activity (S) a minimum convex polygon (a triangle) can be created; its area covers the individual activity space. The minimum convex polygon is the result of connecting all activity locations visited by the individual. Additionally, other geographic measurements can be applied to analyze whether a distribution of point events indicates some directional trend. One of these measures is the standard deviation ellipse (SDE), depicted by the ellipse in Figure 5.9. The SDE is centered on the mean center of all the points (locations) of the activity pattern, here the three locations of home, work and shopping, whereas the long axis of the ellipse is oriented in direction of the maximum dispersion and the short axis in direction of the minimum dispersion (Buliung and Kanaroglou, 2006b, pp. 180-1). In addition to the visual effect, the SDE helps to discover if the revealed behavior follows patterns of the transport supply system or land-use systematically.

Based on the minimum convex polygon, the straight-line distances for each trip leg can be calculated for this particular example (see Figure 5.9). The lengths are 10.6 km for t_1 (HW), 3.7 km for t_2 (WS) and 7.9 km for t_3 (SH). Respectively, the detour factor for this tour is $(t_2+t_3)/t_1$ or 11.6/10.6 = 1.09. The detour factor will in any case be equal to or above 1; only if the secondary activity is located exactly on the straight-line distance between Home and Work the detour factor is equal to 1. Since detour factors are estimated using the EOD survey data, more information is available about modes used, about the socio economic-attributes of the person, about the household a person belongs to and the temporal aspects. For instance, the detour example in Figure 5.9 was realized by mode combination TpTpTp with the work activity starting at 9.05h, lasting until 16.30h and an overall daily travel time being 120 minutes.

Staying with the example, the detour factor of 1.09 could now be used to estimate a search space considering the information about straight-line distances between the TAZ. In Figure 5.10 the TAZ accessible given a detour factor of 1.09 are marked dark grey. This means, dark grey TAZ, i.e. its centroids, lie within a spatial distance that is equal to or less than the distance between home and work multiplied by 1.09. An interesting aspect that requires explanation is why the actually chosen activity location (S) is not part of the search space: this is due to the fact that the detour factors derived from the EOD are based on trips with geographic coordinates on the block level, i.e. on the denominated micro level. The estimation of the search space – considering a detour factor – is based on the TAZ level, i.e. on the macro level. In this example, the detour factor of 1.09 was calculated as mentioned above; the respective detour factor on the TAZ level would have been 1.11 (11.4 km (HW), 4.0 km (WS), 8.6 km (SH)) and would have been based on the straight-line distances between the TAZ centroids. Given the criterion to select only those zones within a detour factor of equal to or less than 1.09, these zones had to be excluded. If we revise the TAZ-centroids of the Home and Work locations in Figure 5.10 (black points), minor differences become evident, as both block-coordinated points are located in such way that they reduce the dimension of the polygon and thus the detour factor. Regarding the estimation of detour factors, this minor aggregation error has to be accepted. Otherwise, we would have to calculate spatial path flows basing entirely on the micro level. As there are more than 50,000 blocks, this would have created a problem of too many spatial options to be handled.

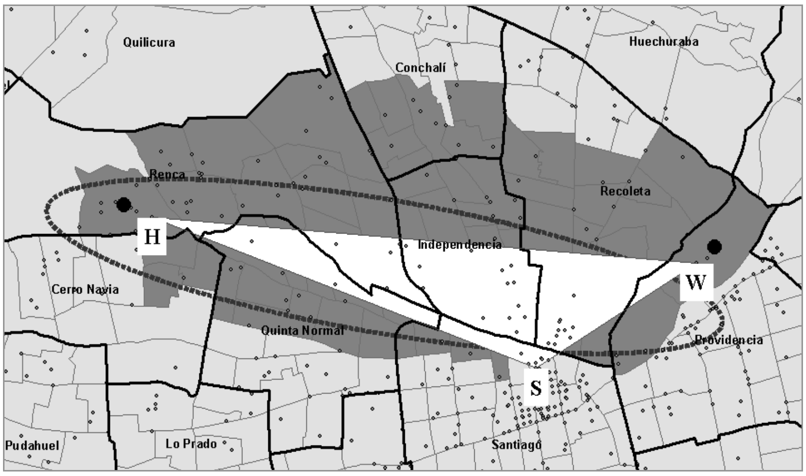


Figure 5.10: Search space given the detour factor

Now, the detour factors can be estimated for each of the EOD activity pattern observations. In addition, we analyze whether they differ according to criteria of socio-economic character or due to temporal and spatial aspects. The objective is to define a set of empirical detour factors that can then be used for an automated process using a GIS for drawing large numbers of search spaces for all Home-Work relations to be considered.

5.4.2 Empirical detour factors

The preceding remarks gave an overview how detour factors can be generated and utilized for the definition of the search space. It may be expected that the detour factors and thus, the number of TAZ selected for a secondary activity, vary according to circumstantial, socio-economic or land-use aspects. Research on urban forms and travel behavior often focuses on identifying and – if possible – quantifying the impact of these factors on travel-related decision-making in the spatial context. For instance, some of the related research considers the following variables: a) socio-economic (age, gender, education), land-use (population density, commerce/service densities), trip costs (Zegras, 2004, testing how these variables influence people's choice to make home-based, non-work, non-school walking trips); b) household structure, income and employment status, number of vehicles available, location variables such as park and ride facilities or distance to the city center (Buliung and Kanaroglou, 2006b, testing the influence of these variables on the household activity space and the daily kilometers traveled); c) area size and travel distance between daily activity locations, building density, retail accessibility, street connectivity and traffic density (Fan and Khattak, 2008, testing the influence of these variables on the individual daily activity space and the daily travel distances). In general terms, these studies focus on identifying urban form conditions that reduce travel (distances) and lower car usage. They are sometimes based on the assumption that dense urban forms reduce the need for travel and thus open possibilities for less transport demand or for more sustainable mode options. Another approach (similar to the one pursued here) is presented by Kim and Kwan (2003), where a potential path area for activities is estimated considering constraints of e.g. maximum travel time thresholds or opening hours of facilities. However, the main distinction is that, given the details considered, the approach remains applicable only to a small number of individuals.

Contrary to the previously named research, here we focus on defining generally valid detour factors that allow in a second step for the estimation of a huge number of search spaces (or potential path areas, as other authors name them). Another distinction, or rather a constraint, lies in the fact that only those explanatory variables are included in the analysis that can later build part of the model. For instance, it might be of interest to investigate to what extent an employment status of an individual influences his or her daily activity space. In the work presented here, for practical reasons it is necessary that any explanatory variable can be identified in the model, so that the empirical findings can be integrated directly into the calculation procedure. This reduces the number of variables that can be considered for analysis to the following: a) the mode combinations considered (see section 5.1); b) the variable of primary activity duration; c) the socio-economic variable, reflected by household user groups, and d) the distance between locations of primary activities, namely home and work locations. The empirical bases for analysis are activity pattern types 'HXYH' and 'HYXH' and detour factors are calculated as illustrated in Figure 5.8. The following tables depict the descriptive statistics of the detour factors for each variable.

Table 5.3: Detour factors by mode combination

Mode Combination	Valid N	Maximum	Mean	Median	Percentile 60	Percentile 70	Percentile 80	Percentile 90
CaTpTp	10*	66,4	22,9	15,8	22,9	33,2	38,2	52,5
CaCaCa	71	60,3	3,2	1,5	1,7	2,1	2,6	5,3
AaAaAa	94	14,1	2,4	1,3	1,6	1,9	3,5	6,1
TpTpTp	297	16,9	2,1	1,3	1,4	1,7	2,5	3,9
AaTpTp	20	23,5	2,7	1,3	1,4	1,9	2,4	4,0
TpAaTp	13*	10,1	2,1	1,3	1,3	1,8	2,1	2,6
TpTpAa	44	62,1	3,9	1,2	1,4	1,6	2,1	3,0
AaCaAa	9*	1,8	1,2	1,2	1,3	1,3	1,3	1,8
AcAcAc	211	41,2	1,9	1,2	1,3	1,5	1,8	2,7
AaTpAa	15*	31,7	3,9	1,1	1,3	1,3	3,9	6,3
TpCaCa	10*	1,7	1,2	1,1	1,2	1,2	1,3	1,5
TpAaAa	42	5,4	1,6	1,1	1,2	1,3	1,5	3,0
TpCaTp	106	29,3	1,4	1,1	1,1	1,1	1,2	1,4
BiBiBi	9*	1,6	1,1	1,0	1,1	1,2	1,2	1,6
TpTpCa	84	3,0	1,1	1,0	1,1	1,1	1,2	1,3

Annotation: ordered by median values, combinations with less than 20 observations are flagged with * / (N=1,035)

Table 5.4: Detour factors by primary activity duration

Primary Activity Duration	Valid N	Maximum	Mean	Median	Percentile 60	Percentile 70	Percentile 80	Percentile 90
3	71	41,2	2,8	1,3	1,5	1,6	1,9	2,8
4	60	219,7	7,9	1,3	1,6	2,1	3,2	8,3
5	98	31,7	2,5	1,3	1,5	1,8	2,5	4,7
6	104	38,7	2,9	1,2	1,4	1,7	2,8	5,5
7	81	11,6	1,9	1,3	1,3	1,5	2,3	3,6
8	101	46,4	2,4	1,2	1,3	1,4	1,8	2,9
9	130	66,4	3,1	1,2	1,3	1,8	3,5	6,2
10	169	11,5	1,6	1,1	1,3	1,4	1,9	2,7
11	179	10,1	1,6	1,1	1,2	1,3	1,7	2,5
12	75	15,3	1,5	1,1	1,1	1,2	1,3	1,8
13	17*	2,4	1,3	1,1	1,1	1,2	1,8	2,1
14	13*	1,7	1,2	1,1	1,2	1,3	1,5	1,6
15	5*	1,8	1,3	1,2	1,2	1,2	1,5	1,8

Annotation: ordered ascending by column ‘Primary Activity Duration’, combinations with less than 20 observations are flagged with * / (N=1,103)

Table 5.5: Detour factors by user group

User Group	Valid N	Maximum	Mean	Median	Percentile 60	Percentile 70	Percentile 80	Percentile 90
no car / <= 150	52	28,7	2,5	1,4	1,6	2,0	3,0	4,6
no car / > 1.200	6*	2,5	1,5	1,3	1,4	1,8	1,8	2,5
1+ cars / <= 150	16*	219,7	16,9	1,3	1,4	2,0	2,2	31,7
1 car / > 1.200 CHP	53	7,2	1,9	1,3	1,4	1,9	2,4	4,1
2+ cars / > 300 & <= 600	29	9,1	1,9	1,3	1,4	1,7	2,7	3,2
2+ cars / > 600 & <= 1.200	47	10,1	2,1	1,3	1,7	1,8	2,1	3,5
1+ cars / > 150 & <= 300	57	37,7	3,8	1,2	1,3	1,8	2,6	10,2
1 car / > 300 & <= 600	165	46,4	2,0	1,2	1,3	1,5	2,1	2,7
no car / > 600 & <= 1.200	64	66,4	3,0	1,2	1,4	1,8	2,5	3,6
2+ cars / > 1.200 CHP	110	41,2	2,1	1,2	1,2	1,4	1,8	2,9
no car / > 300 & <= 600 CHP	201	17,2	1,8	1,2	1,2	1,4	1,7	2,7
1 car / > 600 & <= 1.200	136	62,1	2,6	1,2	1,3	1,5	2,1	4,3
no car / > 150 & <= 300	167	60,3	2,2	1,1	1,3	1,4	1,7	3,5

Annotation: ordered by median, combinations with less than 20 observations are flagged with * / (N=1,103)

Table 5.6: Detour factors by distance between home and work

Distances (in meter)		Valid N	Maximum	Mean	Median	Percentile 60	Percentile 70	Percentile 80	Percentile 90
1	0 to 1,010	102	66,4	7,7	2,0	2,5	4,6	8,6	23,5
2	1,011 to 2,441	103	17,2	3,0	1,5	1,8	2,9	4,7	6,6
3	2,442 to 3,912	102	12,3	2,5	1,6	1,9	2,4	3,3	5,1
4	3,913 to 5,635	103	10,1	2,0	1,3	1,6	2,1	2,7	3,8
5	5,636 to 7,219	101	4,7	1,6	1,2	1,4	1,7	2,1	2,7
6	7,220 to 8,666	102	6,8	1,5	1,2	1,2	1,3	1,6	2,5
7	8,667 to 10,743	103	3,3	1,3	1,1	1,2	1,3	1,5	1,8
8	10,744 to 13,425	102	2,0	1,1	1,0	1,1	1,2	1,3	1,4
9	13,426 to 17,337	103	2,6	1,2	1,1	1,1	1,2	1,3	1,5
10	17,338 to 32,260	102	2,0	1,1	1,0	1,1	1,1	1,1	1,3

Annotation: ordered ascending by column ‘Distances (in meter)’ / (N=1,023)

The column for minimum values is excluded, as these values are equal to 1 in all cases given the definition for the estimation of detour factors. From the descriptive analysis the following observations can be made:

- **General observation:** across all variables significant differences between mean and median values become evident and illustrate the influence of statistical outliers; the descending order of median values often gets mixed-up after percentile 70 (except for the distance categories) and the search for a systematic link between the variables and the descending detour factors becomes more difficult.
- **Mode Combinations:** the median values range between 1.0 (TpTpCa, BiBiBi) and 1.5 (CaCaCa) (the values of CaTpTp seem to be heavily influenced by outliers and are not considered here); the lowest detour factors (1.0 to 1.1) show mode combinations where walking trips build part of the combinations; the medium detour factors (1.2 to 1.3) mainly show combinations where a car is used as well as a pure mode combination of TpTpTp; the outstanding value of the detour factor 1.5 for CaCaCa might be explained by its low total walking distance (median: 1.8 km), which means that origins and destinations of primary activities are within short distance; an increased walking distance in the second or in the last leg of the pattern rapidly increases the detour factor (the same argument applies in principle for CaTpTp); division into different groups of mode combinations is not absolutely unambiguous, as, for instance, the combination of TpAaAa with a detour factor 1.1 is in between those combinations where a walking trip leg seems to be responsible for the low detours; the most evident argument for a differentiation is related to whether or not a walking trip builds part of the combinations; eventually, six of the 15 mode combinations appear in less than 20 observations reducing the statistical validity of these values.
- **Primary activity duration:** the median values range between 1.3 and 1.1; although not in an absolutely linear fashion but it can be observed that the longer the primary activity takes, the smaller the detour factor gets; an exception are hour ‘7’ and hour ‘15’, the value of the latter is not too reliable due to a small number of observations; this tendency persists for the percentile 70 but is heavily stirred in the upper percentiles.
- **User groups:** the median values range between 1.1 and 1.4; there are no systematic differences that can be related to the differentiation of user groups; neither the car availability nor the income distributions indicate a clear pattern of an ascending or descending order of detour factors.
- **Distance:** the median values range between 2.0 and 1.0; similarly to the primary activity duration and not in an absolutely linear fashion, values descend as distance increases, meaning that the longer is the distance between the home and work location, the smaller is the detour factor; this general observation is valid in the higher percentiles.

The descriptive statistics indicated that both the time spent on and the distance to the primary activity (thus, the time spent on the trip from Home to Work given the distance) had a clear influence on the detour value. The significance of the mode combination was less clear, only walking trips seemed to be relevant. No evidence of a systematic influence of the socio-economic household user group variable on the detour factor was observed.

Prior to testing the variables via regression models, a cut-off point for outliers was defined. The detour factors observed in activity pattern types 'HXYH' and 'HYXH' were analyzed by percentiles, starting from percentile 70 upwards. To this end, the detour factor value of one percentile was divided by the factor value of the previous percentile, thus allowing for calculating the percentage increase by each percentile step. The following Figure 5.11 illustrates the percentage increase. The percentile 88 was identified as cut-off point, which is equal to a detour factor of 2.9 (bigger grey point). After this point the almost linear increase was interrupted, and factors speed up in an erratic manner.

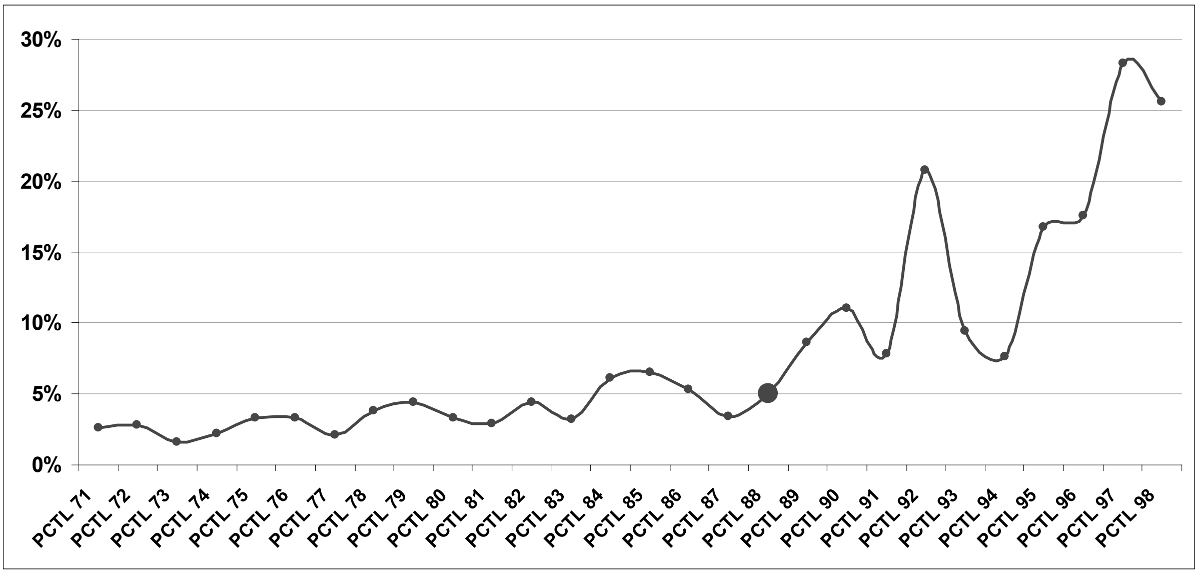


Figure 5.11: Percentage increase of detour factors

Annotation: PCTL = Percentile

For the purpose of testing the influence of variables on the detour factors, several generalized linear regression models were estimated. According to the preliminary conclusions based on the descriptive analysis, the regressions were done against the following background:

- The mode combinations were aggregated into the following five categories, due to the observed similarities and, in particular, due to the low number of observations: 1) pure public transport combination (TpTpTp); 2) pure car travel combinations (AcAcAc, AaAaAa) – the first two categories were analyzed individually due to a higher number of observations available and their relevance as in sum they represent more than 50% of all chosen combinations; 3) combinations with a walking trip between home and work locations (CaCaCa, CaTpTp); 4) mixed combinations of public transport and car passenger trips (AaTpTp, TpAaTp, TpTpAa, AaTpAa, TpAaAa); 5) further combinations including walking trips (AaCaAa, TpCaCa, TpCaTp, TpTpCa, and BiBiBi).
- The durations of primary activities were aggregated into two time periods, following the ascending order of detour factors with increasing duration of activity. The first time period

comprises short durations of less than six hours, the second time period comprises durations of six hours or longer. Splitting into three different time periods (< 6 hours, 6 to < 10 hours and >= 10 hours) was tested as well, but the distinction between the two latter groups did not prove to be statistically significant.

- The previous analysis provided no indication how or if at all to consider the socio-economic user groups. Nonetheless, several regression models were tested considering the user groups in a disaggregated fashion as well as considering a grouping by five income groups and by two car-ownership groups (see Table 4.1). In the final model, the socio-economic variable was not considered.
- The distances were grouped into seven categories (percentiles 10 to 60 and above).

Given the possibility of including up to four variables and the numerous options of grouping the variables, many different models were estimated. At this point, only the final model is shown including the variables of primary activity duration and distance between home and work. The following Table 5.7 shows results of the final estimates.

Table 5.7: Influence of distance and time on detour factors - parameter estimates

Dependent Variable: detour factor					
Parameter		**B**	**Std. Error**	**t-statistics**	**Significance level**
	Intercept	1.134	.0222	51.179	.000
Work duration	less than 6h at work	.097	.0339	2.844	.005
	6h or more at work	0^a	.	.	.
HW distance	<= 1.47 km	.343	.0483	7.111	.000
	> 1.47 km & <= 3.37 km	.268	.0478	5.602	.000
	> 3.37 km & <= 5.15 km	.363	.0479	7.575	.000
	> 5.15 km & <= 6.75 km	.258	.0478	5.400	.000
	> 6.75 km & <= 8.04 km	.201	.0475	4.236	.000
	> 8.04 km & <= 9.51 km	.162	.0475	3.414	.001
	> 9.51 km	0^a	.	.	.

Annotation: 0^a, Set to zero because this parameter is redundant, Goodness of fit: 0.163 / (N= 902)

As mentioned before, several more regressions were carried out including the user and mode combination groups. None of the models including user groups (or their differentiation by income and/or car ownership) allowed for the estimation of statistically significant parameters. In contrast, a significant relationship was found in the mode combinations, especially when walking trips were involved which confirmed the earlier findings of the descriptive statistics. However, in the final regression model the mode combinations were not included as their effect on the detour factors is already captured by the distance variable. This becomes evident if we look at the mode combination of CaCaCa. Detours were highest (and parameters significant) because of the short distance between home and work. Conversely, similar result came up in the analysis of the detour factors by distance, indicating the highest detour factor for the shortest distance between primary activities.

The analysis of the activity duration indicated a dependency between the time spent at work and the range of the detour. A simple linear regression between both variables showed a negative sign, indicating that the longer is the time spent on the primary activity, the more likely is the decrease of the detour factor (beta: -0.127, t-test: -3.8, R^2: 0.13). Accordingly, Table 5.7 shows that the detour factor is likely to increase if less time is spent at work. As mentioned before, first three time periods were considered, but no statistical evidence was found that would have allowed distinguishing between the longer duration time periods. Eventually, the distance variable was included distinguishing between seven distance categories. As the variable was originally divided into ten equally sized distance categories, the relationship between distance and detour factor diminished with a longer distance between home and work locations. As soon as the straight-line distance exceeded 10 km, the reduction effect on detour factors was not observed anymore. Hence, the longer distances percentiles were grouped sequentially until it was found that the differentiation was still significant if the distance was grouped into six percentiles (or up to a maximum distance of 9.5 km) with one category containing any Home to Work relation of longer distances (at maximum 32.3 km). Notice that these thresholds do not coincidence with those shown in Table 5.6, as these were calculated before the application of the cut-off factor.

In summary, the analysis suggests differentiation of detour factors both by duration of the primary activity and by distances between primary locations. In practical terms this means that later in the empirical application these factors determine the geographical magnitude of the search space. The socio-economic user groups did not have a significant influence on the detour factors and are not considered as an attribute when it comes to the application. However, their influence is captured indirectly. Time regimes, i.e. duration of the primary activity, are dependent on the user groups. Now, detour factors are found to be dependent on the duration. Notice that conducting this analysis we attempted to identify the relevant variables that can be reproduced by the model. Without doubt, further variables may influence the detour factors. For example, integration of more disaggregated socio-economic attributes offers a potential for further analysis. It is likely that the consideration of disaggregated activities such as shopping or leisure leads to further detour factors. However, we focus here on the detection and quantification of the general relevance of variables that build part of the model. In this sense, a more disaggregated analysis using more variables probably improves the model's refinement but not necessary its general functionalities.

The following Table 5.8 summarizes values for predicted and observed detour factors. We made predictions applying the regression model – where variables are dummies – to each observation in the data set and then estimating the mean value for each category of work duration and Home-Work (HW) distance. The table shows that the predicted and the estimated mean values are close to each other and are different among categories; this supports the classification process. Based on that we then obtained the values of percentiles 60, 70 and 80 which show the percentages of all trips that have a detour factor equal to or less than the value obtained.

Table 5.8: Detour factors by time periods and distance categories

Work Duration	HW Distance	Valid N	Predicted	Observed				
				Mean	Median	PCTL 60	PCTL 70	PCTL 80
less than 6 hours	<= 1.47 km	39	1,57	1,63	1,48	1,64	2,00	2,10
	> 1.47 km & <= 3.37 km	20	1,50	1,62	1,46	1,70	1,87	2,24
	> 3.37 km & <= 5.15 km	14*	1,59	1,56	1,51	1,62	1,82	1,98
	> 5.15 km & <= 6.75 km	11*	1,49	1,29	1,24	1,26	1,32	1,52
	> 6.75 km & <= 8.04 km	20	1,43	1,38	1,20	1,28	1,44	1,61
	> 8.04 km & <= 9.51 km	20	1,39	1,41	1,24	1,32	1,44	1,68
	> 9.51 km	63	1,23	1,21	1,10	1,16	1,29	1,39
6 to 15 hours	<= 1.47 km	52	1,48	1,44	1,29	1,50	1,65	1,80
	> 1.47 km & <= 3.37 km	70	1,40	1,37	1,20	1,27	1,44	1,64
	> 3.37 km & <= 5.15 km	75	1,50	1,50	1,17	1,38	1,64	2,06
	> 5.15 km & <= 6.75 km	79	1,39	1,42	1,15	1,29	1,44	1,97
	> 6.75 km & <= 8.04 km	71	1,34	1,35	1,15	1,20	1,41	1,76
	> 8.04 km & <= 9.51 km	71	1,30	1,29	1,14	1,19	1,29	1,39
	> 9.51 km	297	1,13	1,14	1,04	1,08	1,14	1,25

Annotation: categories with less than 20 observations are flagged with * / (N=902)

The detour factors of percentile 70 were used for the forthcoming estimation of the geographical search spaces (further explanations to this decision is given in section 6.2.3). In general terms, the data table shows the following pattern: the lowest detour factor appears (1.14) if there is a long-distance trip between home and work and more than six hours were spent at work. On the opposite, the highest detour factor (2.00) is estimated for a short-distance trip between home and work and if only up to six hours were spent at work. This means that values generally tend to decrease within each time period and in case of longer distances, but start already at a lower level for the second – 6 to 15 hours – time period. There are few exceptions to these tendencies but with minor differences between the factors. For instance, in the shorter time period the factors for the first three distance categories are quite similar (between 1.82 and 2.00), then their value drops down to 1.32, nonetheless, a conclusion is difficult to draw as the number of observations in this category is low.

Detour factors and, respectively, their practical use for the calculation of search spaces represent the spatial constraint. With the analysis above we intended to identify a set of detour factors that most likely form search spaces where the majority of travel options is included. The results of this section point out that the respective values of Table 5.8 can be a good 'starting solution' for defining the spatial scope for the search for secondary activities locations. Nevertheless, it still remains to be tested whether the estimated spatial paths – based on the detour factors shown here – reproduce the spatial distribution of empirically observed paths in the EOD. We discuss this issue in the upcoming chapter on the results. The final section introduces several more relevant inputs for the calculation procedure and deals with the aspect of tour relations, i.e. with the question whether the search for secondary activity locations occurs in the proximity to the home or work locations.

5.5 Local attraction

So far, the search space is only a representation of a number of zones in the proximity to the primary activities locations. Considering land-use and travel times needed to reach zones within the search space expands this perspective by bringing in aspects that later allow for the weighting and ranking of choice alternatives. Below we show the data regarding land-use, discuss the issue of tour relations and introduce the travel times (distributions) used both from network model ESTRAUS and from the EOD.

5.5.1 Attractiveness

The zonal attractiveness, i.e., the land-use density (LUD), is derived from land-use model MUSSA and is defined as density of land-use according to the size of zonal area. Land-use model MUSSA provides, inter alia, the built area in square meters by land-use category, i.e. industry, commerce, service, education and other. With the focus on the activity pattern type ‘HXYH’ and specifically on secondary activities of shopping, leisure and other activities (reflected by ‘Y’), the density of commerce and service land-use is considered representative for the land-use attractiveness. The following Figure 5.12 illustrates the attractiveness as a sum of commerce and service-oriented land-use by TAZ.

Land Use Density in Santiago by Traffic Analysis Zones

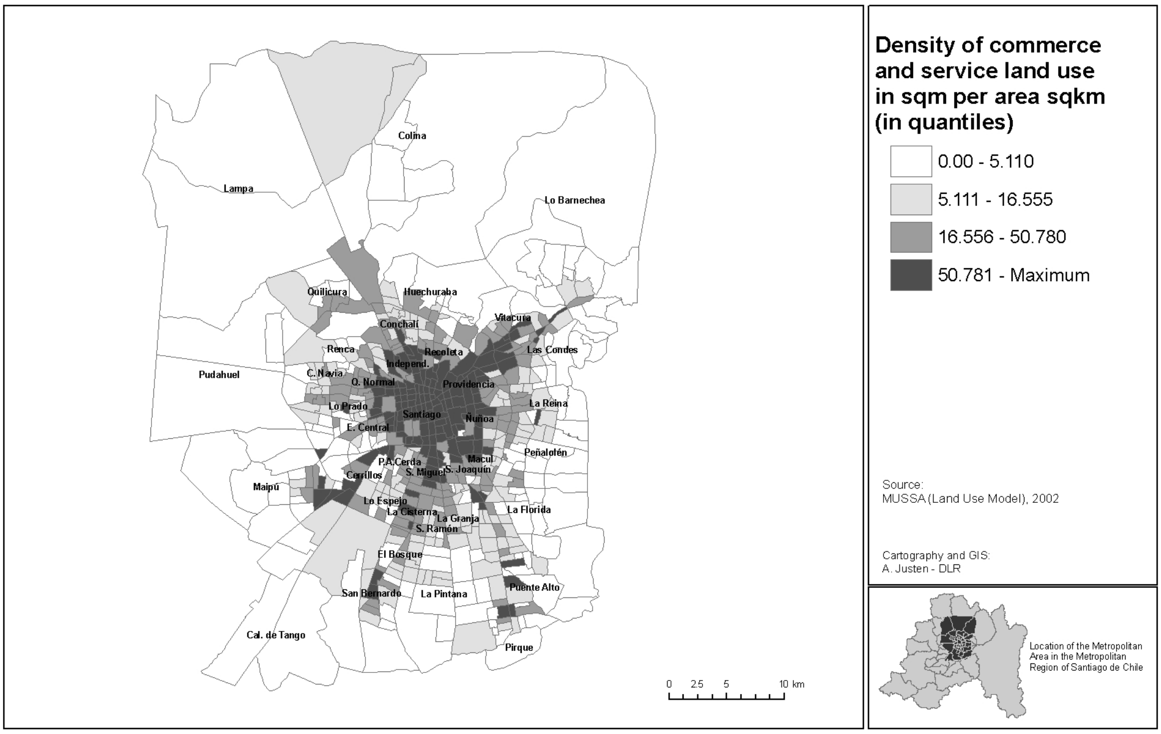


Figure 5.12: Santiago Metropolitan Region – density of commerce and service land-use

Given the land-use of each TAZ, we can estimate the relative attractiveness of any zone within search spaces dividing the single density of a zone by the summed density of a search space. This allows for ranking TAZ related to the search space according to their land-use characteristics (for further explanation, see section 6.2.4).

5.5.2 Tour relations

Distinction between home- and work-related tours is another characteristic of the approach. Generally, a trip to the activity ‘Y’ in the pattern type ‘HXYH’ represents a trip between work and home locations. Nevertheless, only from the tour ‘HXYH’ it is not clear whether the location decision for activity ‘Y’ is made due to the spatial proximity to its trip origin ‘X’ or destination ‘H’, or, in other words, whether the trip was work- or home-related. In case of the mode choice it looks different: the choice of how to travel to activity ‘X’ determines the decision on the adjacent mode choice how to get to activity ‘Y’ and then home to ‘H’. If the location of activity ‘Y’ is in the direct proximity to activity ‘X’, the travel time distributions should be based on the trip leg of ‘XY’ (for the later calibration of the model), as the decision to do activity ‘Y’ has its origin in the activity ‘X’. Conversely, if the location of activity ‘Y’ is in the proximity to home ‘H’, the respective parameters for calibration should be based on ‘YH’. The following Figure 5.13 illustrates this distinction in accordance with the proximity of activity ‘Y’ (Y_1, Y_2) to home ‘H’ and to activity ‘X’.

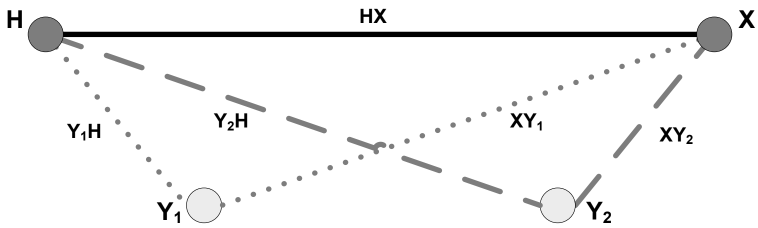


Figure 5.13: Scheme of home- and work-related tours

In the case of Figure 5.13, Y_2 is denominated as work-related, Y_1 as home-related, given the spatial proximities either to activity ‘X’ or home ‘H’. This issue can be further illustrated by an example of a tour with mode combination TpTpCa. If a trip by public transport between home and work is relatively long and adjacent mode decisions are possible, it is likely that a second long public transport trip and a short walking trip follow. This means that the trip from work to the secondary activity ends in a short (walking) distance to the home location. Hence, the decision for the secondary activity location is actually taken in accordance with the home location, although the trip initially started from work. This tour would be denominated as home-related.

Within the empirical application the possibility of both home- and work-related tours is established. Due to the need of a rule to be able to differentiate within the empirical data of the EOD whether the spatial proximity of activity ‘Y’ to home ‘H’ or to activity ‘X’ is given or not, the straight-line distances between the locations were calculated and analyzed. To clearly identify the spatial orientation of each activity pattern in the data set, we established the following rule:

- If the spatial distance between activities ‘X’ and ‘Y’ is longer than between ‘Y’ and ‘H’, the secondary activity location is considered as home-related, otherwise – as work-related.

One may state that this distinction is less relevant in case of a short distance between both locations. If home and work locations are within walking distance, say 500 meters, it is likely that a person sees no difference between work and home-related surroundings. However, for the sake of an unambiguous rule we assume that all tours are either home- or work-related. Nonetheless, it might be an interesting research question, starting from what minimal distance upwards individuals perceive the surroundings of primary locations as independent.

As discussed above, we can further assume that the differentiation between work- and home-related tours depends on the mode combination used – at least for some mode combinations. The example of TpTpCa demonstrated that given a long distance between home and work, it is less likely that only a short trip to an activity located next to the work activity occurs, because this would lead to a very long (in distance and time) walking trip back home. The following Table 5.9 illustrates the share of home- and work-related tours for a set of mode combinations.

Table 5.9: Share of home and work related tours by mode combinations

	home related		**work related**	
	Valid N	in %	Valid N	in %
TpTpTp	66	44.3	83	55.7
AcAcAc	42	45.2	51	54.8
TpCaTp	3	5.3	54	94.7
AaAaAa	13	81.3	3	18.8
CaCaCa	6	50.0	6	50.0
TpTpCa	37	97.4	1	2.6
TpTpAa	15	78.9	4	21.1
TpAaAa	8	50.0	8	50.0

Annotation: Activity pattern type 'HXYH'; mode combinations with less than 10 observations were excluded from the table. For any other mode combination considered in the model, we assume an equal share of home- and work-relations. / (N=400)

Although the analysis is not possible for all mode combinations due to data limitations, the given combinations already yield some important insights. The pure car driver combination (AcAcAc) and the pure public transport combination (TpTpTp) have a slight tendency of being work-related, whereas the car passenger combination indicates a strong tendency towards home-related trips. A very clear orientation is observed for the combinations of TpCaTp and TpTpCa. As expected, in the former case almost 95% of all tours are work-related; in the latter case similar share is home-related due to the final walking trip leg. Although the specific reasons for the observed tendencies remain unexplained (except for TpTpCa and TpCaTp), it can be concluded that the orientation matters for some mode combinations and that the model should be able to reflect these tendencies.

The interrelation between used modes, available daily travel time and the tour relation perspective is not trivial regarding their inclusion in the model. We stay with the example of mode combination TpTpCa and the indicated share of 97.4% home-related versus 2.6% of work-related trips observed in the EOD. The model then has to allow for the estimation of both tour types (home- and work-related), but at the same time it has to reflect that trip flows are primarily home-related. However, in some cases of specific mode combinations, e.g., TpTpCa, the daily travel time thresholds already ensure that choice options in the proximity to the home location are more likely. The practical consideration of the tour relations is described in more detail in the section covering the results, see 6.2.5.

5.5.3 Accessibility

Another important input to the approach are impedances, measured in travel time by mode between all zones of the study area and search spaces, respectively. In the section on the methodology it was shown that travel times are important during different stages of the approach.

a) The sum of the daily travel times over all trip legs has to remain within given thresholds.
b) Dividing the land-use attractiveness by the accessibility in travel time per mode allows ranking choice options related to the search space.
c) The reproduction of the travel time distributions towards secondary locations is objective of calibration of the destination choice.

Naturally, the travel time needed to reach a certain zone within the search space varies due to the characteristics of travel modes, motorized modes being faster and non-motorized modes slower. The importance to include the travel time into the analysis of ranking choice options becomes evident considering different network qualities and associated travel times. Differentiation by mode results in different number of zones a person manages to reach in 10, 20 or 30 minutes of travel. We obtain clear evidence regarding this if we have a look at origins and destinations reachable within time slots. This information can generally be observed in an impedance matrix of the study area provided by the network model. For Santiago the number of origins and destinations reachable within timeslots of 10, 20 and 30 or more minutes by mode are shown in the following Table 5.10. Here, the total of 381.924 OD pairs refers to the 618 Traffic Analysis Zones.

Table 5.10: OD travel times by mode and time slots

	car		public transport		walking		cycling	
	N	%	N	%	N	%	N	%
0 to 10 min	38.155	10,0%	2.741	,7%	1.257	,3%	6.133	1,6%
11 to 20 min	97.774	25,6%	17.014	4,5%	1.893	,5%	15.597	4,1%
21 to 30 min	111.668	29,2%	37.283	9,8%	2.983	,8%	23.949	6,3%
> 30 min	134.327	35,2%	324.886	85,1%	375.791	98,4%	336.245	88,0%
Total	381.924	100%	381.924	100%	381.924	100%	381.924	100%

Annotation: Values taken from ESTRAUS transport network model, 2007. Notice that intrazonal travel times for each TAZ were estimated dividing by 2 the mean value of the three shortest travel times to neighbouring TAZ.

Table 5.10 indicates that network quality and associated connectivity of the number of zones that are reachable within 10, 20 and 30 minutes vary significantly by mode due to distinct mode velocities. The highest number of OD pairs connected within 30 minutes is reflected by private car (64.8%), followed by public transport (14.9%), cycling (12.0%) and walking (1.6%). Within the empirical application travel time matrices of the type shown above are used to estimate a) the total daily travel time, and b) the rank of choice options. Regarding c) and the provision of travel time distributions by mode for destination choice calibration, the respective information is derived from the EOD. The use of modeled times via ESTRAUS for the estimation of spatial paths and their calibration against values of the EOD creates some discrepancies which are subject of a more detailed discussion in section 6.4.

Further on, the objective is to introduce the travel time distributions of the EOD, used as 'target values' during the iterative process of secondary activity destination choice calibration. The respective data analysis considers all trips to locations of secondary activities like shopping or leisure. Given the distinction between home- and work-related tours introduced before, the travel time distributions towards 'Y' (Y_1, Y_2, respectively) are calculated from EOD data. The following Figure 5.14 shows the frequency distributions by grouping trips into time slots of 10 minutes each.

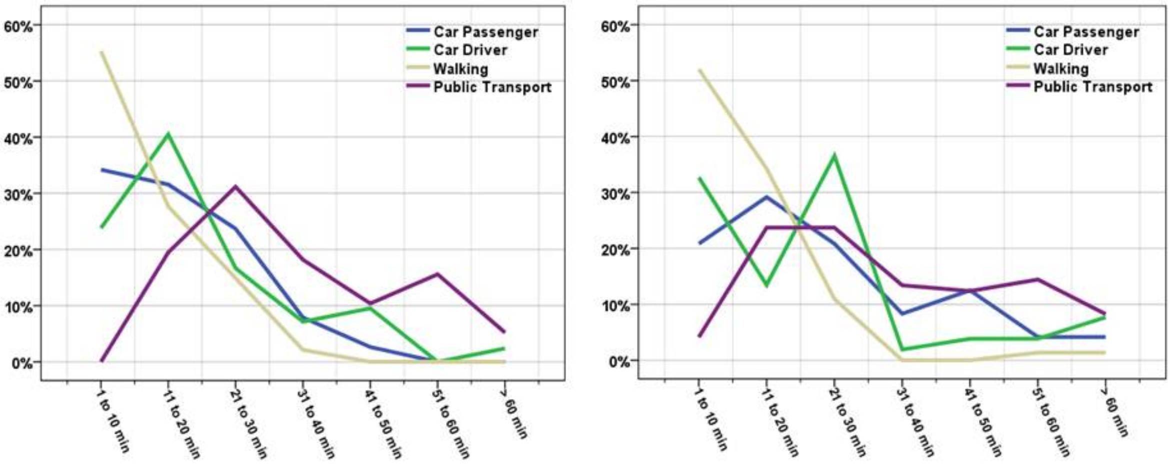


Figure 5.14: EOD travel times by modes and relation to secondary activity locations

Annotation: Activity pattern type 'HXYH' where 'X' was reported as work activity / left: home-related / (N=204); right: work-related / (N=246)

Both distributions are influenced by the fact that respondents tend to round up their real travel times to 5 or 10 minutes intervals. For instance, persons tend to report a real travel time of 23 minutes as of 25- or 30-minutes duration. This explains the 'peak-like' appearance of the distributions. However, strong similarity can be observed for walking trips, which hardly occur above the 30 minutes threshold. Public transport distributions follow similar pattern as well, with some sharper peaks in case of home-related distribution. A difference is observed for the car driver distributions where the majority of home-related trips occur within 20 minutes while work-related trips seem to last a little longer, with the peak at 30 minutes of travel time. This indicates that if a secondary location is chosen next to the home location, it appears to be slightly nearer than if a secondary location is chosen starting from work. The same argument applies for the observed car passenger distributions where work-related trips towards the secondary location appear to be a little longer than in the home-related case.

5.6 Summary

The empirical analysis performed in the previous sections aimed at the identification of appropriate thresholds for a set of time-space constraints. To outline the analytical steps, the activity pattern type ‘HXYH’ was chosen, representing a challenge of modeling conditional decisions of a macro and micro scale regarding activities, times, modes and locations. To begin with, the number and type of the most relevant mode combinations for ‘HXYH’ were determined. Based on 15 mode combinations, the time constraint of the total daily travel time was estimated for each of the observed combinations. For this purpose empirical thresholds were identified allowing for the incorporation of the great majority of the observed behavioral heterogeneity.

In the next step, the analysis focused on the temporal pattern of starting times and duration of the ‘X’ and ‘Y’ activities of activity pattern type ‘HXYH’. In the first descriptive step, the distributions of starting times and duration were revised aiming at an aggregation of the data to the most relevant time periods. The influence of socio-economic and geographical variables on the temporal patterns was analyzed. The driving assumption was that socio-economics or geography may have an influence on the temporal organization of activities. In compliance with the variables actually represented in the model, the user groups and Santiago city sectors were subject of further analysis. Correlation results revealed a significant influence of both variables on the time regimes. The strong correlation between user groups and city sectors allowed for the conclusion that they partially explain the same phenomenon, so we decided to consider the socio-economic user groups as a distinguishing feature of the time regimes.

Subsequently, we introduced the concept of detour factors that represent the spatial constraint for the estimation of search spaces. Regression analysis was conducted in search of relevant aspects that influence the magnitude of detour, namely mode combinations, time spent on the primary activity, socio-economic user groups and distances between the primary locations of home and work. It was found that the duration of the primary activity and the distances between primary locations affect the detour; thus, the empirically derived factors were differentiated by these variables. At the same time, the analysis did not indicate a significant relationship regarding user groups. It was argued that this finding is most likely due to the aggregated fashion of the socio-economic variable and that a more profound analysis of disaggregated socio-economic attributes could offer further insights.

The final section focused on the aspects of attractiveness and accessibility as essential factors for ranking destinations related to the search space. Zonal attractiveness is simply the function of the proportional density of each TAZ on the number of TAZ that conform the search space. The accessibility is measured in travel times from a primary activity location to the secondary activity location. The latter aspect is of crucial importance as distributions of the travel time needed to reach a location within the search space serve as a reference for the calibration of destination choice. In addition, this section introduced the distinction between home- and work-related tours based on the spatial proximity of activity ‘Y’ to home or to work location.

As a result of the previous sections, all parameters necessary to ‘run’ the empirical application have been defined. Notice that these values are result of the respective analysis based on the information available for the city of Santiago. Thus, in a different spatial context the respective empirical values have to be determined according to the local conditions. We summarize the parameters that will be used in the model as follows:

- 15 mode combinations as defined in section 5.1.
- Maximum daily travel times by mode combination as defined in Table 5.1.
- Time regimes (starting times and duration of primary and secondary activities) by user group as defined in section 5.3.1 and 5.3.2.
- Detour factors for the calculation of time-space dependent search spaces as reflected in Table 5.8.
- Land-use attractiveness as illustrated in Figure 5.12, tour relations as depicted in Table 5.9 and accessibility via travel time matrices provided by ESTRAUS.

The upcoming chapter illustrates the main results of the model application considering the identified parameters. In addition, we focus on validating the obtained results against the EOD.

6 Model application and results

The methodology introduced in chapter 2 and the empirical values specified in chapter 5 are now applied for the estimation of spatial path flows. The following overview summarizes (and partially repeats) the main data sources and outlines some important structural elements of the model.

- The study area is divided into 618 TAZ, as shown in Figure 4.1.
- The OD matrices for the home-work relation by mode and user group are obtained from transport model ESTRAUS. Commuting as well as impedance matrices date back to a model of 2007. Travel times for walking and bicycle trips were estimated afterwards assuming an average velocity of 4 km/h (walking) and 12 km/h (bicycle), respectively.
- Conceptually the model considers five modes: car driver, car passenger, public transport, walking and bicycle. As data for bicycle travel behavior in the EOD is very sparse and the transport model ESTRAUS has not considered bicycles as a single mode yet, no spatial paths can be calculated; nevertheless, the approach considers the mode in its design.
- The demand is disaggregated into 13 user groups as defined in Table 4.1.
- The land-use of each TAZ is given by land-use model MUSSA and considers service and commerce activities in square meters. This information goes back to a model of 2002.
- Trip production, i.e. the estimation of pattern demand, is based on the number of user groups by TAZ given by land-use model MUSSA for 2002 and the activity pattern type shares indicated by the EOD.
- Any information concerning empirical thresholds for maximum daily travel times, detour factors and time regimes of starting times and duration are derived from the EOD.

The computational algorithm was implemented employing the software package SPSS using the software programming language (Brosius, 2008). Respectively, all data flows and routines for the estimation of spatial path flows were documented as a comprehensive sequence of commands. In general terms, the programming commands follow the generic model introduced in section 3.2 and illustrated by Figure 3.2. The respective syntax – thus, the entire computational model – was moved to the Digital Annex, see also Annex-Table 7: Overview of SPSS-Syntax to run the model; required input and generated output files are described in Annex-Table 8: Overview of the model's input and output data files.

The first section shows the results of trip production denominated as pattern demand. In result, we obtain the number of conducted 'HWYH' tours by user group and origin. In the following section we explain how search spaces are calculated given the empirical detour factors. In an intermediate validation step we compare the matching levels between EOD and the model regarding the observed secondary locations to be found inside the calculated search spaces. Next, we illustrate the reduction effect on choice options of a limited maximum daily travel time. Subsequently, we describe the processes of weighting, ranking and selection of a reduced choice set. We then introduce the calibration results regarding the adjustment of the secondary activity location choices to empirically observed trip distributions by mode. The final sections deal with a model scenario using adjusted travel times; an illustration of the spatial path flows; a comparison of modeled daily travel times with those observed in the EOD and finally, a quantification of the reduction potential of the time-space constraints.

6.1 Estimation of pattern demand

The pattern demand is defined as a product of the number of households by user group and zone and the probability that an activity pattern is conducted. The number of households by zone is provided by land-use model MUSSA; thus, the activity pattern probability (APP) remains to be defined. With the typology of activity pattern (see Table 4.3) we grouped patterns by a set of criteria that allowed for the estimation of APPs. To this end, the typology accounted for the column headed ‘APT category’ where the activity pattern types (APT) were grouped into 19 further categories. The APPs are now calculated for these categories excluding all patterns with three or more additional stops on a tour. The following Table 6.1 depicts the APPs by 13 household user groups.

Table 6.1: Activity pattern types probabilities by user group

		HOUSEHOLD USER GROUPS (part 1)													
APT category	APT examples	no car / <= 150		1+ cars / <= 150		no car / > 150 & <= 300		1+ cars / > 150 & <= 300		no car / > 300 & <= 600		1 car / > 300 & <= 600		2+ cars / > 300 & <= 600	
		Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %
1	HXY	1038	31.5%	111	27.8%	2590	39.6%	664	33.9%	2763	45.9%	1608	40.7%	194	40.8%
2	HYH	837	25.4%	93	23.3%	1306	20.0%	396	20.2%	1043	17.3%	733	18.5%	80	16.8%
3	HXHYH	264	8.0%	24	6.0%	646	9.9%	195	9.9%	553	9.2%	374	9.5%	55	11.6%
4	HYHYH	476	14.5%	59	14.8%	719	11.0%	231	11.8%	480	8.0%	340	8.6%	38	8.0%
5	HXHYHYH	73	2.2%	10	2.5%	151	2.3%	62	3.2%	122	2.0%	101	2.6%	10	2.1%
6	HYHYHYH	186	5.6%	26	6.5%	334	5.1%	112	5.7%	211	3.5%	130	3.3%	12	2.5%
7	HYHXHYHXH	18	.5%	1	.3%	32	.5%	10	.5%	24	.4%	11	.3%	2	.4%
8	HYHYHYHYH	68	2.1%	11	2.8%	99	1.5%	29	1.5%	60	1.0%	29	.7%	4	.8%
9	HXYH	51	1.5%	12	3.0%	171	2.6%	54	2.8%	224	3.7%	175	4.4%	28	5.9%
	HWYH	**16**	**.5%**	**6**	**1.5%**	**93**	**1.4%**	**20**	**1.0%**	**148**	**2.5%**	**83**	**2.1%**	**10**	**2.1%**
10	HYYH	118	3.6%	18	4.5%	131	2.0%	60	3.1%	138	2.3%	114	2.9%	12	2.5%
11	HXYHYH	30	.9%	12	3.0%	62	.9%	26	1.3%	79	1.3%	70	1.8%	7	1.5%
12	HYHYYH	39	1.2%	6	1.5%	93	1.4%	43	2.2%	70	1.2%	61	1.5%	9	1.9%
13	HXYHYYH	2	.1%	0	.0%	2	.0%	4	.2%	0	.0%	5	.1%	0	.0%
14	HYYHYYH	4	.1%	1	.3%	4	.1%	5	.3%	4	.1%	6	.2%	3	.6%
15	HYHYXHYH	15	.5%	4	1.0%	17	.3%	4	.2%	18	.3%	18	.5%	3	.6%
16	HYHYYHYH	41	1.2%	2	.5%	56	.9%	17	.9%	36	.6%	40	1.0%	6	1.3%
17	HYHXYHYYH	0	.0%	0	.0%	6	.1%	2	.1%	2	.0%	2	.1%	0	.0%
18	HYHYYHYYH	3	.1%	0	.0%	5	.1%	1	.1%	4	.1%	5	.1%	2	.4%
19	HYXYH	31	.9%	9	2.3%	117	1.8%	45	2.3%	191	3.2%	131	3.3%	11	2.3%
	Sum	3294	100%	399	100%	6541	100%	1960	100%	6022	100%	3953	100%	476	100%

		HOUSEHOLD USER GROUPS (part 2)												
APT category	APT examples	no car / > 600 & <= 1.200		1 car / > 600 & <= 1.200		2+ cars / > 600 & <= 1.200		no car / > 1.200		1 car / > 1.200		2+ cars / > 1.200		
		Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Sum
1	HXY	716	48.4%	1037	41.3%	327	39.9%	96	59.6%	389	44.8%	645	44.1%	12178
2	HYH	227	15.4%	457	18.2%	126	15.4%	20	12.4%	122	14.1%	193	13.2%	5633
3	HXHYH	127	8.6%	230	9.2%	86	10.5%	11	6.8%	82	9.4%	133	9.1%	2780
4	HYHYH	110	7.4%	192	7.6%	62	7.6%	6	3.7%	56	6.5%	93	6.4%	2862
5	HXHYHYH	18	1.2%	47	1.9%	12	1.5%	1	.6%	9	1.0%	24	1.6%	640
6	HYHYHYH	38	2.6%	65	2.6%	18	2.2%	3	1.9%	14	1.6%	22	1.5%	1171
7	HYHXHYHXH	2	.1%	6	.2%	4	.5%	0	.0%	1	.1%	2	.1%	113
8	HYHYHYHYH	12	.8%	15	.6%	5	.6%	0	.0%	3	.3%	5	.3%	340
9	HXYH	70	4.7%	152	6.1%	50	6.1%	9	5.6%	61	7.0%	111	7.6%	1168
	HWYH	**47**	**3.2%**	**75**	**3.0%**	**26**	**3.2%**	**6**	**3.7%**	**21**	**2.4%**	**31**	**2.1%**	**582**
10	HYYH	37	2.5%	57	2.3%	25	3.1%	7	4.3%	30	3.5%	35	2.4%	782
11	HXYHYH	24	1.6%	54	2.2%	20	2.4%	1	.6%	23	2.6%	34	2.3%	442
12	HYHYYH	21	1.4%	35	1.4%	12	1.5%	2	1.2%	11	1.3%	33	2.3%	435
13	HXYHYYH	0	.0%	4	.2%	1	.1%	0	.0%	4	.5%	4	.3%	26
14	HYYHYYH	2	.1%	3	.1%	2	.2%	1	.6%	1	.1%	5	.3%	41
15	HYHYXHYH	4	.3%	7	.3%	8	1.0%	0	.0%	4	.5%	14	1.0%	116
16	HYHYYHYH	3	.2%	13	.5%	12	1.5%	0	.0%	6	.7%	9	.6%	241
17	HYHXYHYYH	0	.0%	1	.0%	0	.0%	0	.0%	2	.2%	2	.1%	17
18	HYHYYHYYH	1	.1%	4	.2%	4	.5%	0	.0%	1	.1%	1	.1%	31
19	HYXYH	66	4.5%	131	5.2%	45	5.5%	4	2.5%	49	5.6%	97	6.6%	927
	Sum	1478	100%	2510	100%	819	100%	161	100%	868	100%	1462	100%	29943

Annotation: The included APPs are assumed to represent 100%, thus reflecting the entire pattern variation observed in the EOD. Overall 29,943 activity patterns are considered. This is due to the exclusion of 2,147 patterns (see not considered pattern, Table 4.3, p. 69) and the fact that only patterns realized on a regular workday (Monday-Friday) were included (minus 9,183 patterns). Notice that the Valid N and % for 'HWYH' are already included in those accounted for in 'HXYH'.

For instance, the table indicates that households with a car and a monthly income between 300,000 and 600,000 Chilean Pesos realize a tour of type 'HXYH' with a probability of 4.4% on a regular working day. Notice that 'HXYH' works as an example here and reflects also tours of type 'HYXH'. Given the grouping by pattern types, a further disaggregated approach would require defining how 'HXYH' is divided into tours, for instance, of 'HWSH', 'HELH' and 'HOWH' types, etc.

Remember that the case used for the estimation of spatial paths flows is a tour of type 'HWYH'. Respectively, the APP for this tour-type has to be considered. We now can already estimate the number of home-work trips by multiplying the number of households by user groups with the shares of 'HWYH' as indicated in the table above (see cells in grey). The following Table 6.2 reflects the same shares of 'HWYH' by user group as above but in addition the number of households provided by land-use model MUSSA. The multiplication of both households and APP leads to a total of 26,474 trips over all user groups. This implies that with the model this number of trips from home to work needs to be further distributed to secondary locations of activity ‘Y’. In consequence the number of 26,474 trips serves later as a reference, as we expect to match this number applying the model.

Table 6.2: Trip demand for pattern type ‚HWYH’

	no car / <= 150		1+ cars / <= 150		no car / > 150 & <= 300		1+ cars / > 150 & <= 300		no car / > 300 & <= 600		1 car / > 300 & <= 600		2+ cars / > 300 & <= 600	
	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %
HWYH	**16**	**.5%**	**6**	**1.5%**	**93**	**1.4%**	**20**	**1.0%**	**148**	**2.5%**	**83**	**2.1%**	**10**	**2.1%**
N UG		**249,930**		**336,366**		**260,733**		**21,531**		**25,764**		**104,211**		**176,868**
Trips		**1,214**		**5,058**		**3,707**		**220**		**633**		**2,188**		**3,716**
	no car / > 600 & <= 1.200		1 car / > 600 & <= 1.200		2+ cars / > 600 & <= 1.200		no car / > 1.200		1 car / > 1.200		2+ cars / > 1.200			
	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %	Valid N	in %		
HWYH	**47**	**3.2%**	**75**	**3.0%**	**26**	**3.2%**	**6**	**3.7%**	**21**	**2.4%**	**31**	**2.1%**		
N UG		**65,047**		**120,511**		**38,651**		**8,397**		**50,278**		**61,887**		
Trips		**2,068**		**3,601**		**1,227**		**313**		**1,216**		**1,312**		

Annotation: Pattern type ‚HWYH’ implies the realization of three trips (HW, WY, YH). Thus the overall number of trips for this pattern type is 26,474 * 3 = 79,422. N UG describes the number of households by user group.

6.2 The hierarchical choice process

In reference to the generic step-by-step approach described in section 3.2, the selection of a set of secondary activity locations, denominated also as alternatives, follows the principles of a hierarchical choice process. The following sections illustrate how a final choice set of alternatives is step-by-step identified and spatial path flows calculated for the entire city of Santiago.

6.2.1 Defining the search space

The delimitation of the search space is essential for reducing the set of potential location options. It was argued that this is rational, given the individual’s limited resources and restricted cognitive capacities to search over the entire space. Applying the detour factors identified in the empirical analysis on the EOD data, the search space is estimated for each home-work relation and is represented by an ellipse. Remember that the size of the search space depends on the time spent on the primary activity, i.e. at work, and on the straight-line distance between home and work. For every OD relation at a TAZ level two search spaces (ellipses) are estimated, one for each corresponding work duration time period (see Table 5.8). Search spaces were created using a Geographical Information System (GIS). For this purpose, the following input information for each OD relation beside the empirical detour factors was required (see also Figure 6.1):

- the XY coordinates of the midpoint (MP) of the home-work axis (HW);
- the extension of the horizontal axis of the ellipse, delimited by S_{max} and E_{max}, calculated by multiplying the (straight-line) distance of HW by the respective detour factor;
- the extension of the vertical axis of the ellipse originating at MP, denoted as $V_{max}MP$, is calculated as $\sqrt{(V_{max}E_{max})^2 - (E_{max}MP)^2}$;

- the rotation angle (RA) for each HW relation assuming a straight-line between both locations.

The following Figure 6.1 illustrates the information required for a calculation of the ellipses, i.e. of the search spaces for each OD relation.

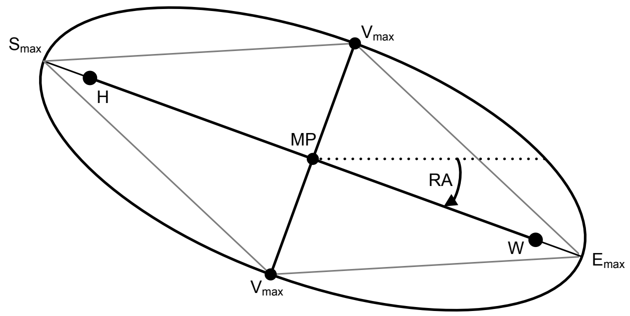


Figure 6.1: Schematic search space ellipse

We illustrate the following calculation example: a distance between H and W of 23.8 km (straight-line), detour factors are 1.29 (if staying at work less than six hours) and 1.14 (if staying six or more hours at work). Respectively, the maximum extensions of the horizontal axes are:

23.8 km * 1.29 = 30.7 km and 23.8 km * 1.14 = 27.1 km

The length of the vertical axes is then defined as:

$$\sqrt{(30.7/2)^2 - (23.8/2)^2} = 9.7 \text{ and } \sqrt{(27.1/2)^2 - (23.8/2)^2} = 6.5$$

Formula 1: Calculation of ellipse vertical axis

The following Figure 6.2 visualizes the calculations examples of search spaces for two home-work relations. The inner (or, respectively, the outer), the smaller (respectively, the larger) ellipses and the black points, i.e., TAZ that fall inside represent zones available for conducting a secondary activity given a stay at work of six or more hours (or less than six hours, respectively).

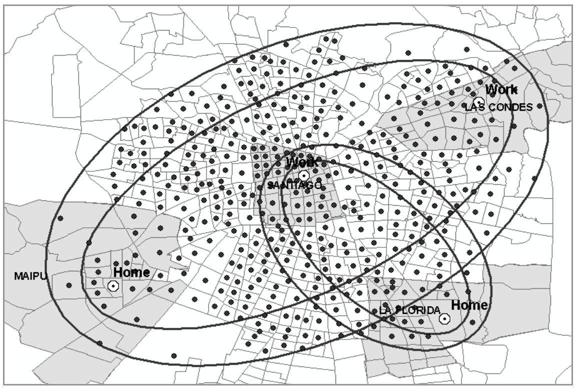


Figure 6.2: Search space ellipses – spatial constraint

Annotation: The white doted points indicate the home and work locations. The grey areas represent all TAZ of the municipalities the home and work locations are situated in (Home locations in Maipú and La Florida, respectively; Work locations in Santiago and Las Condes).

The same calculation was conducted for all home-work relations considering the two working time periods. In accordance with the 29,305 relations from home to work indicated exogenously by transport model ESTRAUS, a total of 58,610 ellipses was calculated (separately for three detour factor sets, according to the respective percentiles of 60, 70 and 80, see Table 5.8). Each ellipse provides a set of potential secondary activity locations. Notice that any alteration of the detour factors naturally alters the extension of the ellipses and thus the potential number of secondary locations. A change in the detour factors might be due to new empirical evidences, for instance, given a different spatial context, i.e. another city. From methodological point of view there is no limitation to include further attributes for defining search spaces via the estimation of ellipses. The limiting factor is then primarily set by the attributes that are represented in the actual calculation procedure.

We know that all zone centroids falling inside one of the ellipses represent a potential secondary activity location. Regarding this, Figure 6.2 indicates an important characteristic of the approach: whether a zone is inside or outside the ellipse depends also on the geographical position of the centroids and, respectively, on the size of the TAZ. For instance, if a share of TAZ surface spatially belongs to the search space but as the centroids are not captured, these TAZ are not selected. This inaccuracy should be recognized but it is rather a characteristic of the underlying zoning system than of the methodology.

6.2.2 The daily travel time restriction

In the next step the search spaces are further refined taking into consideration the time constraint that is the maximum daily travel time by mode combination. Given an OD relation and a search space this time constraint implies that these secondary activities are excluded from the alternatives that are not accessible due to the empirically defined thresholds. To outline this effect, the following Figure 6.3 shows the reduction of selected secondary activity locations according to the time constraint.

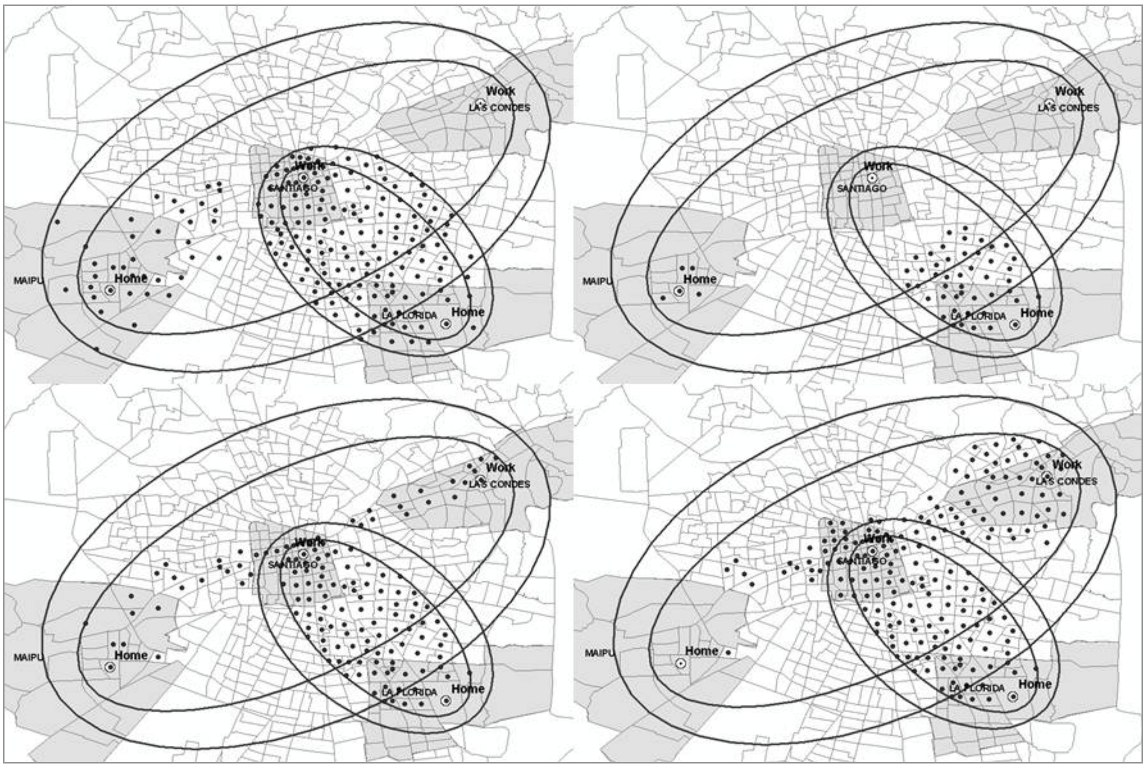


Figure 6.3: Search space ellipses – time and spatial constraints

Annotation: upper left: TpAaTp, TP = 1, upper right: TpTpCa, TP = 2, bottom left: AaTpTp, TP = 2, bottom right: TpAaAa, TP = 2 (TP = Time Period with 1 = less than six hours at work, 2 = six or more hours at work)

Some of the mode combinations have an important effect on the number of feasible secondary locations. For instance, in the case of the TpTpCa combination (upper right) the long distance between home and work locations results in that only very few secondary locations in the direct proximity to the home location may be reached given the maximum travel time available (in this case, 180 minutes). The fact that only locations next to the home location are feasible seems reasonable given that the final trip leg is realized walking. In general, the same effect can be observed for the ellipses of the home-work relation in shorter distance. In this case, more locations remain feasible but tend to be situated closer to the home location. The comparison between the bottom left and the bottom right figures reveals the following: The AaTpTp combination (bottom left) reveals that given the two public transport trips, the degree of freedom is reduced. Potential secondary locations seem to be strongly attached to Santiago's main public transport axis, the 'Alameda'; nonetheless, options in the close proximity to the home location remain accessible, too. In case of a long distance between home and work the TpAaAa combination (bottom right) offers more 'spatial flexibility', given the trip towards the secondary location is done by car. However, options next to the home location require more travel time and are skipped, possibly due to the reduced accessibility of locations in this area (longer travel times).

The figures also show that only an additional reduction has been achieved so far while no differentiation has been made among the remaining alternatives in the ellipse. All alternatives that form part of the search space and accomplish the time-space constraints are equally feasible. Hence, in a following step the results of the processes of weighting, ranking and selection of secondary alternatives are exemplified. But before an intermediate validation of the effects of the time-space constraints is possible. To this end, the number of secondary locations reported in the EOD that fall inside the estimated search spaces is analyzed.

6.2.3 Matching level: search spaces and EOD

The definition of search spaces allows comparing the matching level, i.e. the number of secondary locations reported in the EOD (of pattern type 'HWYH') that were captured by the estimated search spaces. Given that the size of the search space depends on the values of the detour, we can assume that a larger definition of the search space leads to a better match with the EOD. However, the extension of the search space proceeds at the cost of the model's efficiency since more potential secondary locations are selected and need to be processed. The following Table 6.3 shows the matching level between locations within estimated search spaces for different detour factors, i.e., for search space sizes.

Table 6.3: Level of matching: EOD vs. modeled search spaces

<table>
<tr><th></th><th>Car Passenger</th><th>Car Driver</th><th>Public Transport</th><th>Walking</th><th>Sum</th><th>Average number of TAZ per search space</th></tr>
<tr><td>EOD</td><td>32</td><td>56</td><td>86</td><td>67</td><td>241</td><td></td></tr>
<tr><td>Search Space based on percentile 60 - Valid N -</td><td>17</td><td>31</td><td>40</td><td>55</td><td>143</td><td rowspan="2">68</td></tr>
<tr><td>Matching Level</td><td>53.1%</td><td>55.4%</td><td>46.5%</td><td>82.1%</td><td>59.3%</td></tr>
<tr><td>Search Space based on percentile 70 - Valid N -</td><td>22</td><td>34</td><td>49</td><td>61</td><td>166</td><td rowspan="2">94</td></tr>
<tr><td>Matching Level</td><td>68.8%</td><td>60.7%</td><td>57.0%</td><td>91.0%</td><td>68.9%</td></tr>
<tr><td>Search Space based on percentile 80 - Valid N -</td><td>22</td><td>41</td><td>56</td><td>63</td><td>182</td><td rowspan="2">118</td></tr>
<tr><td>Matching Level</td><td>68.8%</td><td>73.2%</td><td>65.1%</td><td>94.0%</td><td>75.5%</td></tr>
</table>

Annotation: 241 activity patterns of ‚HWYH' type were selected from the EOD for the analysis. For comparison three different search space settings were selected based on the detour factors of percentiles 60, 70 and 80 (see Table 5.8). No differentiation is made by time spent on the primary activity, as too few cases for comparison remained.

The comparison of matching levels is differentiated by modes used to reach the secondary activity location. For instance, in case of public transport and search spaces based on detour factors of percentile 70, 57% of secondary locations observed in the EOD are captured by the estimated search spaces. Generally, the results reveal that the highest matching levels are obtained for walking trips and the lowest for trips by public transport. Higher detour factors, i.e. larger search spaces, lead to an improved overall matching level. The increase of 9.6% (6.6%) from percentiles 60 to 70 (and to 80, respectively) shows that matching benefits decrease, while enlarging of the set of potential secondary locations results in on average 38% and, respectively, 26% increase of TAZ per search space. This latter observation is crucial for a compromise between a high matching level and a reasonable reduction of choice options, which has to be reached to keep the upcoming search process efficient.

The dependencies of search spaces on the macro locations 'H' and 'X' and on the time spent on activity 'X' (almost 70% for percentile 70) support the basic assumption of a hierarchical choice process, and that locations of activities 'H' and 'X' determine the location of activity 'Y'. Regarding the relevant differences in matching levels between modes, it is assumed that this is caused, at least partially, by the fact that search spaces were estimated based on straight-line distances rather than on actual network connections. It is likely that this deviation is even stronger for public transport trips since the network supply is concentrated on the specific routes (e.g., bus or metro lines), whereas roads are distributed more evenly. This effect can be observed in Table 6.3 with the lowest matching levels for public transport trips. Additionally, since walking trips generally cover short distances, the estimated search spaces more likely draw the decisions taken in reality. If motorized modes are used (a car or public transport), the likelihood of longer distances between home and work increases and the estimated search spaces are less likely to reflect the actually taken decisions. However, further analysis is required to clarify whether the remaining non-matching 30% cases are because of a mismatch result of considering straight-line distances or due to a pure random effect.

6.2.4 Ranking of secondary activities locations

The ranking of TAZ pertaining to the search space is done according to the criteria of land-use attractiveness and the travel time by mode needed to reach a location. First, the land-use attractiveness (LUA) is defined as a measurement of density in each TAZ. The LUA are calculated using the output of Santiago's land-use model MUSSA. Figure 6.4 indicates the shared attractiveness of each TAZ on the search space based only on its land-use. The relative attractiveness is reflected by the different dot sizes summing up to 1 (or 100%) for each search space.

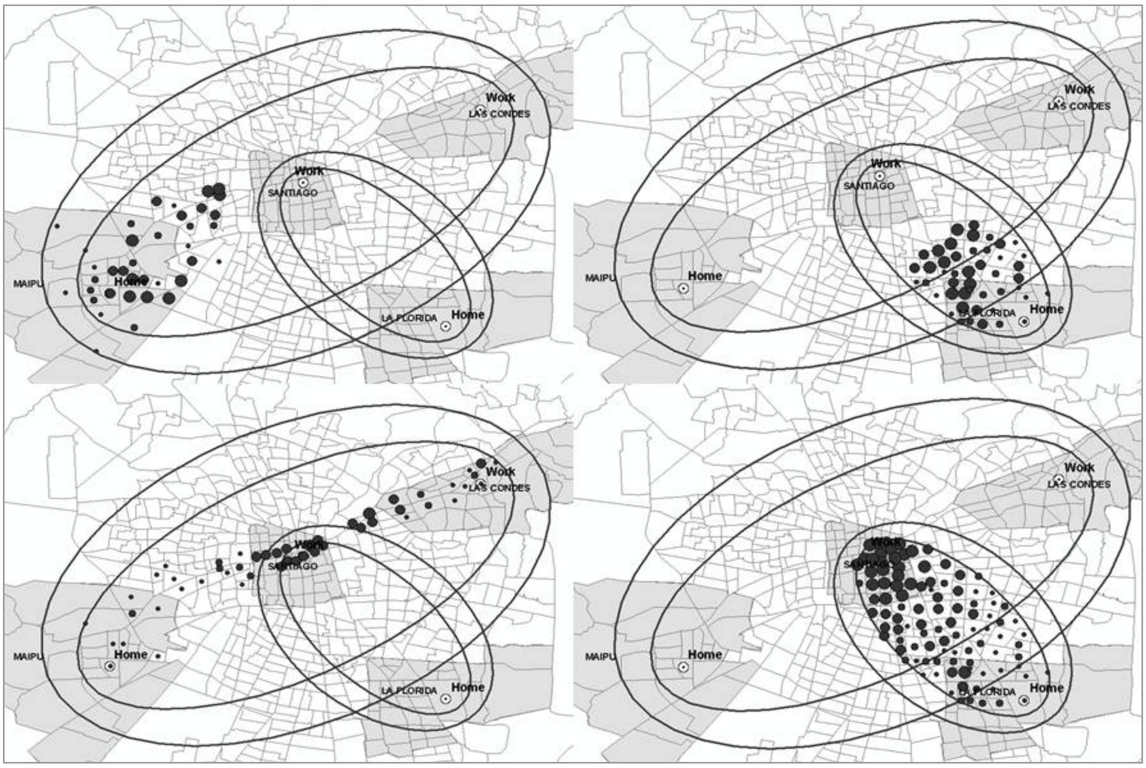


Figure 6.4: Search space ellipses – land-use attractiveness

Annotation: upper left: TpAaTp, TP = 1, UG = 10; upper right: TpTpCa, TP = 2, UG = 5; bottom left: AaTpTp, TP = 2, UG = 8; bottom right: TpAaAa, TP = 2, UG = 4 (TP = Time Period, UG = User Group)

The next step is to define a method to model destination choices within search spaces. For this purpose a procedure for estimating local attraction factors for each TAZ is defined; then, these options are ranked, and a final choice set of secondary alternatives is chosen. The local attraction factor (LAF) is defined combining land-use index LUA and travel times by mode towards secondary locations. Therefore, the LUA_{is} (defined for each TAZ *i* of the search space *s*) is divided by the travel time of the specific mode *m* and travel time category *d*, $TT_{m,d}$):

$$LAF_{is} = \frac{LUA_{is}}{TT_{m,d}}$$

Formula 2: Local attraction factor

Next, TAZ are ranked by LAF values within the search space; this generates an ordered set of location options within each search space. Notice that the LAF are ranked by mode-travel time category *m,d*, for correcting the LAF-based order where the majority of the selected alternatives would be represented by zones with high land-use attractiveness within short travel time distance – a bias compared to observations (refer to Annex-Table 5 for the observed distributions of trips towards secondary activity locations by mode combinations and tour relation). However, the ranking via the LAF is only a mean to define a further reduced choice set taking into consideration Miller's (1956) argument that an individual is able to process only a limited number of options.

6.2.5 The final choice set

Prior to the selection of the final choice set (FCS), we introduce a random process. The attractiveness of TAZ solely has so far depended on the land-use, meaning that the TAZ with the highest land-use attract the highest share of trip flows. Without doubt, the attractiveness of each alternative is not sufficiently described by its land-use value only. We may assume that further attributes contribute to the attractiveness (e.g., quality of services/products offered, kindness of staff, proximity to other facilities, etc.), which have not been not considered here since we have included only the built area of commerce and service activities. We may conclude that further aspects – invisible to the model – influence the decisions taken by individuals, i.e. by user group, for or against a location. This leads to a certain level of uncertainty when it comes to the selection of choice alternatives. This uncertainty is included through the application of a random process: all alternatives, i.e. their LUA, are several times multiplied (concretely 25 times in the application) by a random factor yielding a value between 0.5 and 1.5 (see also 3.2). As a result, the zone with the highest land-use is not automatically the most 'attractive' zone. Random values between 0.5 and 1.5 assure that at maximum the LUA is halved or gain 50% of its original value. Eventually, all TAZ are multiplied by the most frequently occurring random factor. Based on the 'new' LUA values now including a process that represents the uncertainty related to individual decision-making, the order defined by the LAF changes and represents the final order of ranked choice options.

According to Miller (1956) we define seven options per primary activity location as the maximum choice set. These options entering the choice set are selected under three criteria:

a) We choose seven options according to the highest ranked alternatives by LAF order for each mode-travel time category *m,d*.
b) Within this selection, we consider the observed distribution of trips by travel time category, so that the percentage of options in the choice set replicates the observed percentage of trips by travel time category in the EOD.
c) We include the aspect of the tours' relation, i.e. the question whether activity 'Y' is situated close to the home or work locations. As theoretically each tour can be home- or work-related, seven options are defined, each in accordance with the primary locations of home and work.

The latter criterion requires some additional explanation. We saw in Table 5.9 the share of home- and work-related tours for a set of mode combinations as revealed in the EOD. For instance, in the case of TpTpCa the EOD reveals that only 2.6% of all tours are work-related. Regarding the application, this implies that for this mode combination a work-related tour should be feasible, but respective trip flows should not cover a share of more than 2.6%. With the consideration of the upper named criterion c) this is ensured. We assume an individual (or in our case a user group) is capable of processing seven alternatives given he or she is located in one of the primary locations, i.e. at home or at work. During the step of estimating the choice-related probability (taken the example of TpTpCa) we normalize the before calculated local attraction factors for the final choice set options against the empirical shares by tour relations, in this case in such way that 2.6% of the path probability is realized by work-related tours and 97.4% by home-related tours (see Table 5.9).

The following Table 6.4 exemplifies the definition of the choice set according to the observed percentage of trips by travel time category in the EOD.

Table 6.4: Example for the definition of the number of choice alternatives

7 Alternatives		**TpTpTp**	
		44%	***56%***
Travel Time Cat.	**in %**	**home related**	**work related**
0 to 10 min	**2.3**	**0.2**	**0.2**
11 to 20 min	**21.8**	**1.5**	**1.5**
21 to 30 min	**27.0**	**1.9**	**1.9**
31 to 40 min	**15.5**	**1.1**	**1.1**
41 to 50 min	**11.5**	**0.8**	**0.8**
51 to 60 min	**14.9**	**1.0**	**1.0**
> 60 min	**6.9**	**0.5**	**0.5**

Table 6.4 shows the distribution of public transport trips towards secondary locations by time category. The percentages are multiplied by the number of alternatives, here seven, to generate the number of choice options entering the final choice set. Notice that this process is done both for the home- and work-related tours. A problem occurs with the unambiguous selection of alternatives due to rounding. For instance, the first category (home-related, 0 to 10 min, 0.2) would not lead to the selection of an alternative out of the ranked options (7 * 2.3% = 0.16, resulting in that no alternative is selected). In contrast, the last category (home-related, > 60 min, 0.5) would contribute one alternative. In the application, we deal with this rounding problem sometimes including even categories with shares below 0.5. The comprehensive overview of the definition of the final choice sets by mode combination and tour relation can be looked up in Annex-Table 5.

The definition of the FCS by time categories does not indicate yet the final relevance of the alternatives for the calculation of spatial path flows. So far, it only follows the objective to provide more alternatives if reality, i.e. EOD observations, indicates a concentration of trips by travel time categories. For instance, we may assume that three alternatives are defined for a time slot of 1 to 10 minutes and only one for the slot of 11 to 20 minutes. This does not imply that as a result 75% of the spatial path flow is represented by the alternatives of the first time slot and 25% by the alternatives of the second time slot. This is due to that not for any search space (and primary location) the total number of seven alternatives is found. For instance, in case of a very small search space it is unlikely that e.g. a trip towards the secondary activity takes more than 20 minutes.

The probabilities of alternatives have so far been based on the LAF as described above. Now, two more steps are necessary before the actual probability of alternatives pertaining to the choice set can be estimated:

- First, as mentioned above we normalize the LAF of all seven alternatives by the tour relation shares as shown in Table 5.9. Regarding the example above of mode combination

TpTpTp, this means that the seven alternatives representing the home-related ‘perspective’ will cover not more than 44% of the overall flow, while the work-related ‘perspective’, i.e. the seven alternatives will cover the rest 56%. In between the seven alternatives we still take care of the underlying attractiveness described by the land-use. This means that the normalization process does not alter the ranking (the relevance) of alternatives constituting either the home- or the work-related final choice set.

- The second process concerns calibration to assure that the spatial path flows reproduce the empirically observed distributions of trips. The calibration of secondary destination choice is described in more detail in the upcoming section 6.4.

Figure 6.5 shows the final choice sets for the examples used throughout this chapter. One exception is the bottom right figure where we show the choice set for mode combination AcAcAc to include one ‘pure’ mode combination beside the mixed mode combinations. Dot sizes represent probabilities that an alternative is chosen. The summed probability over all alternatives is 1 (or 100%) and shares between the home- and work-related alternatives are adjusted according to their shares by mode combinations (see Table 5.9). Alternatives are ranked by relation and in descending order of their probability, meaning that H1 (W1) represents the home (work) related alternative with the highest probability, H2 (W2) with the second highest, and so on.

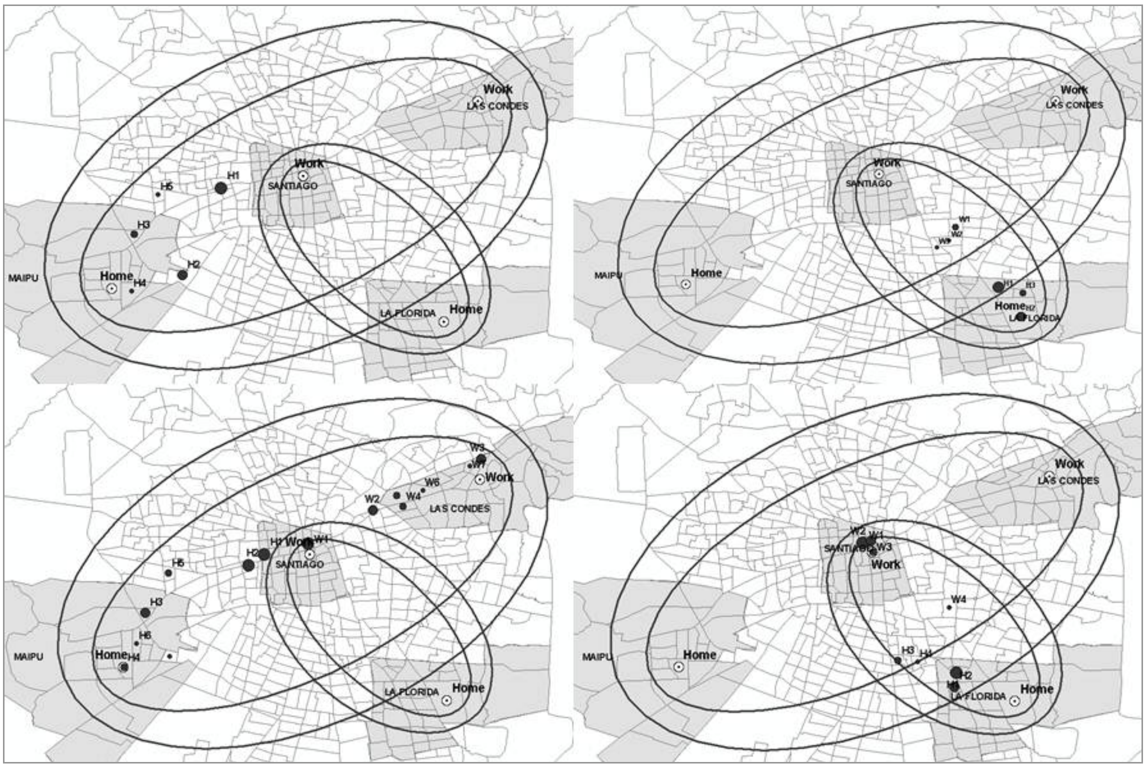


Figure 6.5: Search space ellipses – final choice sets

Annotation: upper left: TpAaTp, TP = 2, UG = 10; upper right: TpTpCa, TP = 2, UG = 5; bottom left: AaTpTp, TP = 2, UG = 8; bottom right: AcAcAc, TP = 2, UG = 13 (TP = Time Period, UG = User Group)

The illustrations allow for further insights of how the introduced mechanisms of a hierarchical choice process actually work, as, for instance, not in every case exactly 14 alternatives (seven for each primary location) constitute the final choice set. Following, we discuss briefly the particularities of the final choice sets according to the related mode combinations.

TpAaTp (upper left): The explanation why five alternatives (home-related only) constitute the final choice set is as follows: the daily travel time limit for this mode combination is 141 minutes and the first public transport trip for a relatively long distance between home and work (23.8 km, straight-line) consumes already 76 minutes. Respectively, only 65 minutes remain for two trips to the secondary activity and back home. This explains why the secondary location – in this specific case – is chosen exclusively in proximity to the home location since the second trip is done by the comparatively fast mode ‘car passenger’ and the major distance of the return trip is covered. In other cases where the distance between home and work is smaller, it is likely that alternatives can be assigned for both cases of a home- and a work-related tour.

The remaining alternatives for the secondary activity are ‘bunched’ together in at maximum 35 minutes travel-time distance to home location (see also Figure 6.4, upper left). Given that the home-related trip is done by public transport and a widely empirical distribution indicating that trips to secondary locations take up to 60 minutes (see Figure 5.14), the model ‘searches’ for potential alternatives over all the time slots considered. But as already indicated, feasible alternatives are to be found at maximum within 35 minutes travel-time distance. According to the selection of the ‘best’ alternatives per time slot, only five (instead of seven) alternatives are selected as shown in the figure. As no work-related alternatives are feasible, the probabilities associated to the five home-related alternatives represent 1 (or 100%).

This example reveals the following: the target of both home- and work-related tours can be achieved not for every OD relation. However, this is a reasonable ‘reaction’ in this specific case onto the application of a time constraint on the daily travel time available. In this example, the temporal constraint determines the geographical locations of the final choice set. However, only five instead of seven home-related alternatives were chosen. This was due to the fact that no options in further travel-time distance could be found.

TpTpCa (upper right): Essentially similar mechanisms are responsible for the reduced number of three alternatives for each tour relation. The differences in dot size between home- and work-related tours indicate that it is very unlikely that the secondary activity location is chosen in the proximity to the work location. Additionally, the limited travel time available (180 minutes) influences the tendency for the secondary location to be home-related. However, seven alternatives were expected to appear in the proximity to the home location. But a more detailed analysis revealed that no alternative was found within 1- to 10-minute walking distance (see Annex-Table 5). Although the EOD observations suggest that four of seven alternatives should be localized in at maximum 10-minute distance, in practice no alternative satisfied this criterion in this specific case. Even the estimated intrazonal walking time outranged the 1 to 10-minute time slot with an average duration of 10.9 minutes. A closer look at the underlying zoning system provides an answer. The relatively large zone of the home location lets us conclude that neither the home zone nor any neighbouring zone alternatives are reachable within 10 or fewer minutes walk. Nonetheless, the remaining alternatives create quite a reasonable result that the home zone obtains the second largest probability.

This example lets us record that the size of the zoning system has an effect on the choice process. Especially the combination of a large zone for the primary activity (either home or work) and a walking trip to the secondary activity will result in no alternatives selected. This is not that much an issue of the methodology but more due to the irregularity of the zoning system. The solution could be either the introduction of a more disaggregated zoning system at the cost of increasing number of zones to be considered or the integration of a heuristic rule always to include the primary location when a walking trip is realized.

AaTpTp (bottom left): In this case, exactly seven alternatives are found for each primary location. This example provides some interesting insights into the underlying land-use and the transport networks in Santiago. All alternatives appear to be 'stringed' along an approximate straight-line between the home and work locations. This seems rational considering the constraints that affect the choice set for this mode combination. The maximum daily travel time threshold is 140 minutes; the distance between home and work is relatively long; and two public transport trips are conducted. Feasible alternatives are closely attached to Santiago's main axis of the 'Alameda' where rapid public transit (bus or metro) is possible. In addition, highest probabilities are reflected by choice options in the proximity to the center of Santiago where the land-use concentration is the highest.

AcAcAc (bottom right): In this case in sum eight alternatives are determined, four of them are home-related and other four are work-related. Using the car for the entire tour allows reaching all potential secondary locations within the daily travel time limit of 140 minutes (not shown). To be more specific, the longest trip (in time) to a secondary location takes only 13 minutes according to the used ESTRAUS travel time matrix. This indicates that in the case of car use the time constraint has no reductive power on the choice set, its reduction is solely result of the spatial constraint. As any location within the ellipse is feasible from the beginning, the final choice set is influenced by the factors of land-use attractiveness and by the 'obligation' of complying with the criteria of a defined number of alternatives by time categories and at the same time reproducing the share of home and work relations for this tour. As no alternatives are found within travel time distance further than 13 minutes, the final choice set is reduced to four alternatives per relation, as additional alternatives simply do not meet the EOD request that also options in further distance should form part of the choice set.

The defined alternatives reflect the impact of very high land-use attractiveness for both home- and work-related tours. While in the city center two zones (W1, W2) indicate comparable levels of land-use attractiveness, the case of the alternative H1 is worth a closer look. Among the home-related secondary locations this zone is by far the most attractive regarding the concentration of commerce and service-oriented activities. The map shows that a large shopping mall ('Mall Plaza Vespucio') is located within this zone.

In summary, several constraints and criteria were applied for the definition of the final choice set; these are a) the spatial constraint via ellipses, b) the time constraint via the daily travel time, c) the land-use attractiveness, and d) a rule-based selection process considering the land-use, the travel time categories and the empirically observed shares of tour relations. The examples and explanations to them confirm the sensitiveness of the approach to these constraints. Several more issues came up indicating space for improvement and refinement of the current setting. For instance, one could consider iterating the process of defining the final choice set to assure that in any case seven alternatives per primary location are estimated. In this connection there is a problem that sometimes no alternatives are found within longer distances which makes it impossible to meet the EOD distributions of trips by mode and time category. However, another solution could be to consider the distance between the primary locations when defining the EOD distributions. In this sense it seems rational that given a short distance between home and work, often a comparatively short trip (in time and distance) is realized to the secondary location. As long as we consider EOD trip distributions equally relevant for the entire city and independent of the distance between home and work, this remains an issue. However, given the current empirical nature of the approach a further differentiation of the EOD is limited as numbers of observations become statistically insufficient.

6.3 The spatial path flow

Before the calculation of the LAF, each zone of the search space was equally likely to be chosen. The respective expression in probabilities meant that each zone indicated the same probability share summing up to 1 (or 100%) over all zones. With the introduction of the LAF, now it was possible to establish a ranked order, which afterwards served as basis for the selection of a reduced set of options by tour relation. In a next step, the shares underlying the final choice set of at maximum seven alternatives per relation were normalized against the share of tour relations by mode combination. In case that no shares could be estimated from the EOD due to a limited number of observations, an equal share of 50% both for home- and work-related tours was assumed (refer to Table 5.9). With this normalization the ‘function’ of the alternatives belonging to the FCS becomes evident: any trip flow from home to work can now be split according to the probabilities of the choice alternatives.

Once the probabilities of the alternatives belonging to the FCS are estimated, the spatial path flows can be calculated. This manifests the multiplicative property of the approach. We use the example of a ‘pure’ public transport mode combination TpTpTp to demonstrate how the spatial path flow is estimated. The following Table 6.5 exemplifies in a reduced form – skipping some auxiliary information obtained during the calculations – the general outcome of the model. Notice that Origin (411) and Destination (255) represent exemplary zones of the home and work locations. The table is split into two parts to improve readability.

Table 6.5: Exemplary results of the model

UG	N_{UG}	APP	PD	Origin (Home)	Destination (Work)	D_W	S_W	D_Y	$PrHW_{Tp}$	Detour Factor	Pr_{TpTpTp}
5	124	2.5%	3	411	255	11	11	1	0.00506	1.14	0.562

Destination ('Y')	$REL_{HR, WR}$	RANK	Pr_Y	$\Sigma PrY_{HR, WR}$	LWP	SPF	TT_Sec (in min)	DTT (in min)
542	1	1	0.21588	0.44	1.00	0.001843	21	102
544		2	0.07834		1.00	0.000669	20	107
404		3	0.05742		1.00	0.000490	12	103
393		4	0.04789		1.00	0.000409	12	99
558		5	0.04048		1.00	0.000346	35	143
246	2	1	0.26271	0.56	1.00	0.002242	18	106
237		2	0.22198		1.00	0.001895	17	106
266		3	0.03308		1.00	0.000282	20	112
273		4	0.02727		1.00	0.000233	21	111
385		5	0.01496		1.00	0.000128	31	103

The table presents the following: 124 households (N_{UG}) belonging to user group 5 (UG) select with a 2.5% probability an activity pattern of type 'HWYH'. This leads to an average pattern demand of 3 (PD). The probability of leaving TAZ 411 (Origin, Home) towards a location in TAZ 255 (Destination, Work) using public transport is defined by $PrHW_{Tp}$ and is given by ESTRAUS. The probability that the public transport trip to work is followed by two subsequent public transport trips is reflected by column Pr_{TpTpTp} and result of the EOD. In addition the time regimes shown in columns represent D_W (duration in hours of the primary activity, 10 to 11 hours), S_W (starting hour of the primary activity, between 10.01h and 11h) and D_Y (duration of the secondary activity in hours).

The feasible TAZ for conducting activity 'Y' are shown in column Destination 'Y'. Thus, this column represents the FCS by tour relation. Column 'RANK' indicates the position of each TAZ within the FCS according to its probability (Pr_Y). The probability of visiting one of these secondary locations (Pr_Y) is result of the application of time-space constraints and the definition of the final choice set as explained in the previous section. Notice that the sum of the Pr_Y by tour relation (see column $REL_{HR,WR}$, with 1 = home-related, 2 = work-related) are normalized against the empirically observed shares (see column $\Sigma PrY_{HR, WR}$). The spatial path flows (SPF) in the case of TpTpTp are then defined as follows:

$$SPF = PD * \Pr HW_{Tp} * \Pr_{TpTpTp} * \Pr_{Y / \sum \Pr Y_{HR,WR}}$$

Formula 3: Calculation of the spatial path flow, example TpTpTp

Additional information representing an outcome of the calculation algorithm is the travel time needed to reach the secondary activity location (TT_Sec) and the expected total daily travel time (DTT). According to the temporal constraint in the case of TpTpTp, DTT does not exceed the maximum limit of 215 minutes. In fact, the estimated DTT remain well below the threshold (99.4 to 112 minutes). However, this appears reasonable given that the Origin-Destination combination reflects the relatively small ellipses in the examples given above. According to D_W the detour factors were calculated and thus the sizes of ellipses, i.e., the search spaces, were defined. Once D_W is assigned, S_W and D_Y can be defined as well (see 5.3). S_W and D_Y are not yet relevant for the estimation of the SPF, but the consideration of a starting time of both the primary and secondary activity opens up a possibility to include time-of-day dependent impedance matrices.

Table 6.5 already includes the location weighting parameter (LWP). The LWP is used for adjusting the estimated travel flows to secondary locations against the empirical observations. Basically, these parameters are used to shift travel flows towards secondary activity locations between modes in case that the modeled differ from the observed distributions. In the previous table this value is set to its default value of 1. The results of the adjustment and calibration process are described in the subsequent section.

6.4 Calibration of destination choice

The calibration of the destination choice for secondary activities is an attempt to reproduce the empirically observed trip distributions by mode. The analysis of the EOD led to the distributions of trips by mode, time slots and tour relation (see Figure 5.14). The parameters applied for calibration were introduced in the section on the methodology and are defined as location weighting parameters. The LWP are multiplied (in each iteration) by the local attraction factor (LAF) to 'shift' probabilities between choice alternatives. Virtually, the LWPs represent parameters applied in a deterrence function in trip-based models. Instead of using a function, the distribution of trips is generated directly from data. In order to simplify the calibration process, the reference trip distribution for calibration is based only on the total number of trips to secondary locations, ignoring tour relation. This reduces the number of parameters to handle and appears reasonable given the relatively small differences between the trip distributions of home and work relations (see Figure 5.14). The following Table 6.6 illustrates the EOD-based joint distribution of home and work related trips to secondary locations by mode.

Table 6.6: EOD results – trip distributions by mode to secondary activity locations

	Car Passenger		**Car Driver**		**Walking**		**Public Transport**	
	Σ Trips to 'Y'	in %	Σ Trips to 'Y'	in %	Σ Trips to 'Y'	in %	Σ Trips to 'Y'	in %
1 to 10min	18	29.0	27	28.7	64	53.3	4	2.3
11 to 20min	19	30.6	24	25.5	38	31.7	38	21.8
21 to 30min	14	22.6	26	27.7	15	12.5	47	27.0
31 to 40min	5	8.1	4	4.3	1	0.8	27	15.5
41 to 50min	4	6.5	6	6.4	0	0.0	20	11.5
51 to 60min	1	1.6	2	2.1	1	0.8	26	14.9
> 60min	1	1.6	5	5.3	1	0.8	12	6.9
Σ	62	100	94	100	120	100	174	100

Annotation: Activity pattern type 'HXYH' where 'X' was reported as work activity. / (N=450)

The LWPs are differentiated by mode and time slot. According to the table above, 28 parameters (4 modes, 7 time slots) have to be defined. The initial calculation is done with the default value of 1 for each time slot and mode combination. In result, the following probability distributions of trips to secondary locations by mode were obtained. Notice that the following Table 6.7 is the result of the estimation of spatial path flows for all 618 TAZ of Santiago, thus the entire city.

Table 6.7: Model-1 results – trip distributions by mode, LWPs = 1

	Car Passenger			Car Driver			Walking			Public Transport		
	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP
1 to 10min	2946	69.9	1	5110	77.1	1	2699	41.5	1	346	3.7	1
11 to 20min	951	22.6	1	1326	20.0	1	2150	33.1	1	4807	52.0	1
21 to 30min	238	5.6	1	171	2.6	1	1652	25.4	1	2644	28.6	1
31 to 40min	52	1.2	1	20	.3	1	0	0.0	1	864	9.3	1
41 to 50min	25	.6	1	2	.0	1	0	0.0	1	380	4.1	1
51 to 60min	0	.0	1	0	.0	1	0	0.0	1	155	1.7	1
> 60min	0	.0	1	0	.0	1	0	0.0	1	50	.5	1
Σ	4213	100		6630	100		6501	100		9246	100	

Annotation: In total, 26,590 trips are distributed to secondary activities by mode. This almost coincides with the initially estimated demand of 26,474 home-to-work trips for this pattern type, see Table 6.2, p. 105) – the minor differences are due to rounding errors.

This first output already provides some important insights as to how the time-space constraints and the definition of the final choice set affect the distribution of trips in the model. We can observe a further concentration of trips by car in the short time distances (compared to Table 6.6), especially for car-driver trips. No secondary locations were found for these modes in distances larger than 50 minutes of travel time and only few within the time slots from 21 to 50 minutes. The same applies for walking trips, but is much more reasonable as trips beyond 30 minutes of duration were rarely observed in the EOD. Regarding public transport, secondary locations are found within even larger distances (time slot of more than 60 minutes travel). Most trips in this uncalibrated model run take 11 to 30 minutes travel time. One can make a preliminary conclusion that many secondary locations concentrate within short travel time distance, specifically, in case of car trips; or, putting it the other way round, too few alternatives remain in larger distance given the constraints applied. Comparing the distributions of 'car passenger' and 'car driver' modes, this effect is even enforced for the 'pure' car-mode combination AcAcAc. The reason is that the 'car driver' mode is not mixed with any other mode combination, while 'car passenger' trips occur in different mode combinations, e.g., in TpTpAa or TpAaAa. In case of these two mode combinations, the daily travel time constraint is less restricted than for AcAcAc. We can now calculate errors E_m for each mode comparing EOD (Table 6.6) with the model (Table 6.7) that is defined as:

$$E_m = \sqrt{\sum_t (V_{m,t}^{Model} - V_{m,t}^{EOD})^2}$$

Formula 4: Error E_m between model and EOD

Annotation: *V* represents the shares of trips towards secondary activity locations by time slot (either by empirical observations of the EOD or by modeled values), *m* represents the modes, *t* the time slots; Error E_m is differentiated by modes *Tp* (public transport), *Aa* (car passenger), *Ac* (car driver) and *Ca* (walking).

The aggregated error *E* (sum over all E_m) between model and EOD, given the default values of 1 for the LWPs, is 155 (E_{Tp} = 35, E_{Ca} = 18, E_{Ac} = 56 and E_{Aa} = 46). Now, we adjust the LWPs manually to 'shift' trip probabilities between the time slots of each mode. After several iterations with the objective of reducing errors E_m the following trip distributions were obtained.

Table 6.8: Model-1 results – trip distributions by mode, LWPs after iterations

	Car Passenger			Car Driver			Walking			Public Transport		
	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP
1 to 10min	1358	35.1	0.001	3165	47.7	0.001	3451	52.1	1.4	328	3.5	0.1
11 to 20min	1217	31.5	0.1	2682	40.5	0.05	2186	33.0	0.85	2744	29.0	0.1
21 to 30min	867	22.4	10	704	10.6	10	993	15.0	0.4	2361	25.0	0.3
31 to 40min	277	7.2	50	72	1.1	50	0	0.0	1	1456	15.4	1.1
41 to 50min	151	3.9	100	8	.1	100	0	0.0	1	1145	12.1	5
51 to 60min	0	.0	1	0	.0	1	0	0.0	1	1072	11.3	50
> 60min	0	.0	1	0	.0	1	0	0.0	1	354	3.7	100
Σ	3870	100		6630	100		6629	100		9460	100	

Annotation: LWPs are obtained by iteratively adjusting the respective parameters. Currently, no automated iteration is possible; thus, results were obtained by manual adjustment.

The indicated LWPs confirm that it is only partially possible to shift trip probabilities for car passenger and car driver modes towards larger time slots. Even drastically small values for the LWPs of up to 0.001 (car driver and car passenger, 1- to 10-minute time slot) are not sufficient to match trip shares to those observed in the EOD. On the other hand, public transport and walking trips could be adjusted more successfully. Overall error *E* for Table 6.8 is reduced to 50, reflecting a good match for public transport (E_{Tp} = 9), walking trips (E_{Ca} = 3) and car passenger mode (E_{Aa} = 7). The major share of *E* is covered by the car driver mode (E_{Ac} = 31). The following Figure 6.6 visualizes the differences between the EOD and the model results differentiated by mode.

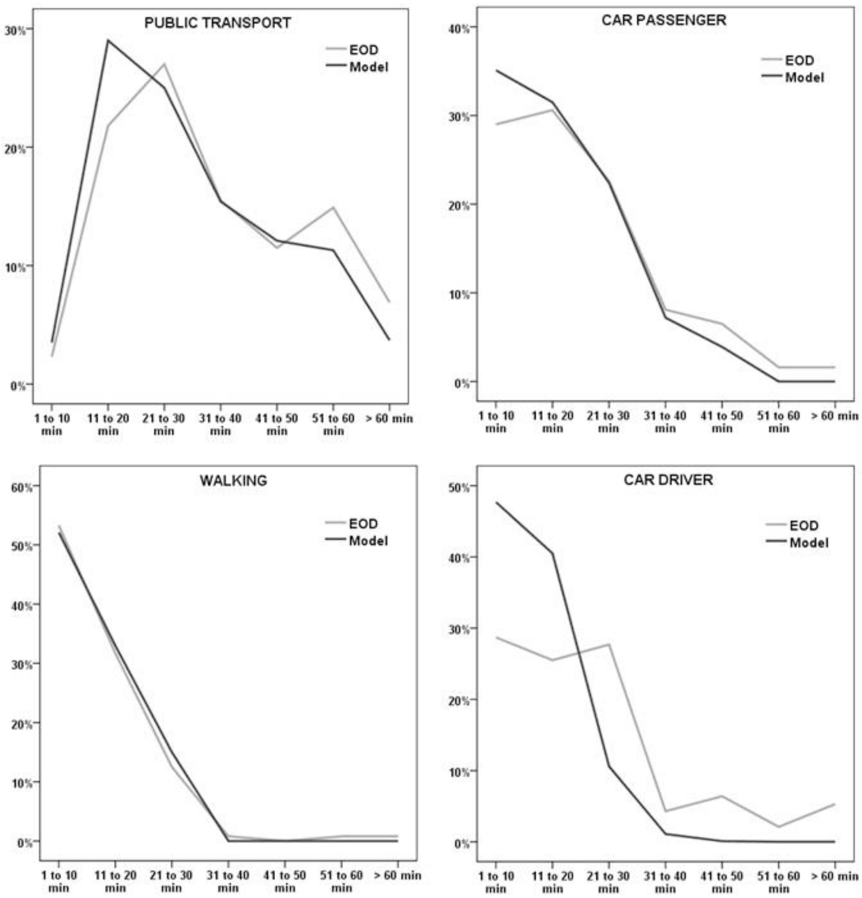


Figure 6.6: Secondary location choice, comparison of EOD and model-1 by mode

The figures clearly indicate that the major problem regarding calibration can be observed for car trips in short time ranges. The differences between EOD and model distributions of car driver mode are above 20% (above 10% for car passenger mode) for the first time slot even after strongly forcing to shift trips using very low LWPs. The reason behind may be that ellipses determining the search spaces are too tight including many alternatives reachable within very short travel time. However, this seems unlikely since the considered detour factors for search space estimation had values for the percentile 70 and allowed for the creation of relatively large search spaces. Another reason may be that car travel times provided by ESTRAUS were too short; as a result, the alternatives within the search spaces are reached faster than trip distributions in the EOD suggests. The effect of possibly too fast car travel times provided by ESTRAUS level-of-service matrices already came up when we defined the final choice set of alternatives (see Figure 6.5).

To test this assumption, we compare the reported travel times by mode of the EOD and those provided by ESTRAUS. For this purpose, we assign every trip pair, i.e. OD relation of the EOD, to TAZ using the respective coordinates. This leads to some inaccuracy, as survey trips were geo-coded on residential block level, whereas ESTRAUS measured travel times are based on TAZ centroids. However, we may neglect this effect for the purpose of this analysis (eventually the large number of survey observations minimizes this inaccuracy as we only look at aggregate average values by mode and time slots). The respective comparison is shown in the following Table 6.9.

Table 6.9: Comparison of travel times – EOD and ESTRAUS

	N	N 'faster'	% 'faster'	0 to 10 min	11 to 20 min	21 to 30 min	> 30 min	N 'slower'
Car	28,285	25,996	91.9%	52.3%	29.0%	10.8%	7.9%	2,289
Public Transport	37,737	28,240	74.8%	34.3%	28.5%	17.1%	20.1%	9,497
Walking	51,408	19,211	37.4%	75.8%	17.2%	5.0%	2.0%	32,197

Annotation: In total, 66,022 motorized trips (Car, Public Transport) and 51,408 non-motorized trips (Walking) of the EOD are compared with respective OD-based level-of-services from transport model ESTRAUS. Columns N 'faster' (% 'faster') represent the number (percentage) of OD relations where ESTRAUS travel times had shorter travel times than those observed in the EOD.

Indeed, the comparison reveals that ESTRAUS network model seems to underestimate car travel times. In over 90% of the observations, the EOD indicates slower travel times. For instance, 52.3% of all relations being faster according to ESTRAUS lie between 0 to 10 minutes of travel time. Similar but less drastic phenomenon is observed for public transport trips. There, almost 75% of all OD-related travel times are slower in the EOD than indicated by ESTRAUS. However, we should point out that the survey was held in 2001, whereas ESTRAUS impedances reflect the network supply quality of 2007. Meanwhile, large-scale road infrastructure projects were finished in Santiago; particularly, the highway ring road was completed. Even considering these factors, the underestimation by the network model appears systematic. The phenomenon of shorter ESTRAUS travel times does not refer to walking trips. There, in only 37% of observations, ESTRAUS produces faster trips than reported in the survey.

This deviation explains – at least for car-based trips – why a great majority of secondary activities locations are found within short travel time ranges to either home or work locations. We may assume that given somehow slower car travel, more location alternatives in further travel time ranges enter the choice set and ease the reproduction of the empirically observed distributions. However, the adjustment of car travel times is somehow difficult. There is disagreement in measurement (residential blocks for EOD, TAZ level for ESTRAUS) and in time when measurement took place (EOD, 2001; ESTRAUS, 2007). In addition, Santiago suffers from an often highly congested morning rush hour. This makes car travel times vary a lot during the day. Taking into account the expansion of road capacities between 2001 and 2007, it seems that some car travel time savings, at least during off-peak hours, are realistic.

To investigate the effect of increased, i.e. 'corrected', travel times on the choice set of secondary activities location, the model is run again with changed travel times. We define this as a 'scenario'. The obtained results are summarized in the upcoming section.

6.4.1 Scenario: Adjusted travel times

We adjust the ESTRAUS impedance matrices to the observed times in the EOD in accordance with the degree of matrices' deviation. Notice that any other parameter (time regimes, thresholds of daily travel time by mode combination, detour factors, i.e. search spaces) remains unchanged in the scenario. Hence, the effect of different choice sets will be solely result of adjusted travel times. However, to avoid an ad hoc alteration of the ESTRAUS matrices, the travel time differences between EOD and ESTRAUS are analyzed in more detail. Therefore, EOD trips and travel times by mode were combined with the respective matrix-based values provided by ESTRAUS. For the purpose of generating adjustment factors, the following steps were conducted.

- Adjustment factors were estimated by dividing the EOD travel times by mode by ESTRAUS travel times by mode.
- All ESTRAUS travel times were assigned to time slots of 10 minutes each, up to a maximum of 60 minutes travel time; all times above 60 minutes were aggregated into one category.

The output is then defined by crossing ESTRAUS travel time categories by mode and estimated adjustment factors. The following Table 6.10 shows the adjustment factors per mode and time slot (median values).

Table 6.10: Travel time adjustment factors

	Car		**Public Transport**	
	Median	Valid N	Median	Valid N
0-10 min	2.96	15,653	2.81	1,526
11-20 min	1.78	7,402	1.61	6,692
21-30 min	1.49	3,278	1.36	8,525
31-40 min	1.34	1,262	1.28	7,844
41-50 min	1.20	500	1.24	5,614
51-60 min	1.10	148	1.18	3,588
> 60 min	.76	42	1.12	3,863

Annotation: We exclude an adjustment of walking trips. The previous comparison (see Table 6.9) does not support the assumption that walking trips are systematically underestimated.

The table reveals that ESTRAUS network matrices did not always underestimate travel times. Factors are highest for car travel and decrease within travel time categories. Indeed, it appears that particularly short car and public transport travel times are strongly underestimated by ESTRAUS since the adjustment in this category will be the highest with a factor of nearly 3. In consequence, the adjustment will be similar for public transport trips but at slightly lower levels. With increased travel times, both the observed and the predicted times approximate. However, the revealed deviations in the observed and the measured travel times should be subject of a more detailed revision, for instance by running the network models of car and public transport at different times-of-day and comparing results with the respective time-of-day dependent values revealed in the EOD. Additionally, it may be tested whether or not the spatial effect of comparing measurements of residential block and TAZ level introduce some bias in the results. Since this was not possible during this work, we accept the generated adjustment factors as such. The obtained factors (by mode and by time slot) were then multiplied by ESTRAUS travel times by mode and time slot. For instance, an ESTRAUS travel time of 23 minutes by public transport is multiplied by factor 1.36 resulting in a ‘new’ travel time of 31 minutes.

Again, the initial calculation for the scenario was done with the default value of 1 for all LWPs. The following probability distributions of trips to secondary locations by mode were obtained with the adjusted travel times for car and public transport included (see Table 6.11).

Table 6.11: Model-2 results – trip distributions by mode, LWPs = 1

	Car Passenger			Car Driver			Walking			Public Transport		
	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP
1 to 10min	1321	33.2	1	2575	38.8	1	2558	41.7	1	40	.4	1
11 to 20min	1351	34.0	1	2150	32.4	1	2038	33.2	1	2715	28.1	1
21 to 30min	1099	27.6	1	1649	24.9	1	1541	25.1	1	3900	40.3	1
31 to 40min	156	3.9	1	224	3.4	1	0	0.0	1	1674	17.3	1
41 to 50min	53	1.3	1	32	.5	1	0	0.0	1	815	8.4	1
51 to 60min	0	.0	1	0	.0	1	0	0.0	1	384	4.0	1
> 60min	0	.0	1	.	.0	1	0	0.0	1	140	1.4	1
Σ	3980	100		6630	100		6137	100		9669	100	

Once again, we calculate error E to compare model and EOD (see Table 6.6). Given the default values of 1 for the LWPs, overall E is 62 (E_{Tp} = 20, E_{Ca} = 17, E_{Aa} = 10 and E_{Ac} = 15, in comparison to 155 considering original ESTRAUS travel times). With the adjusted travel times, particularly, for car travel, more options were entered into the final choice sets in the range of 11 to 30 minutes travel time. This 'eases' the calibration process and makes the model and EOD match better already in the uncalibrated version. This result appears reasonable given that the final choice sets, i.e. the seven alternatives per primary location, are selected in such way that the percentage of choice options reproduces the observed percentage of trips by travel time category in the EOD.

After testing different sets of LWPs, we obtained the trip distributions that reflect an improved match, i.e. a low value of 19 for error E (see Table 6.12).

Table 6.12: Model-2 results – trip distributions by mode, LWPs after iterations

	Car Passenger			Car Driver			Walking			Public Transport		
	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP	Σ Trips to 'Y'	in %	LWP
1 to 10min	1195	30.3	0.5	2126	32.1	0.35	3300	53.4	1.5	80	.8	3
11 to 20min	1314	33.4	0.6	1605	24.2	0.4	1993	32.3	0.85	2118	21.9	0.5
21 to 30min	842	21.4	0.5	2213	33.4	1	883	14.3	0.4	2649	27.4	0.5
31 to 40min	368	9.3	3	607	9.2	5	0	0.0	1	1581	16.3	0.9
41 to 50min	219	5.6	10	79	1.2	10	0	0.0	1	1132	11.7	2
51 to 60min	0	.0	1.00	0	.0	1	0	0.0	1	1442	14.9	10
> 60min	0	.0	1.00	.	.0	1	0	0.0	1	670	6.9	20
Σ	3938	100		6630	100		6175	100		9672	100	

Annotation: The calibrated scenario results represent the final model and the respective data file was entered into the Digital Annex, Annex-Table 8: Overview of the model's input and output data files.

As expected, we are able to obtain a good match between model and EOD applying much more 'relaxed' LWPs. Errors E by mode are low with E_{Tp} = 2, E_{Ca} = 2, E_{Aa} = 4 and E_{Ac} = 11. We conclude that the adjustment of travel times – at least for car travel – is imperative to assure that within the search spaces locations are found in realistically time ranges of up to 40 minutes travel time. Again, it is recommended to inspect network-model-based travel times in comparison to observed times in the survey. Another option is to differentiate EOD trip distributions by distance categories between home and work locations. For the moment, we assume one distribution valid for the entire city, independent of the size of the search space, thus of the distance between home and work locations. If the home-work relation is short in distance, it is likely that different trip time distributions are valid than in the case of a large distance between primary locations. There is further research needed to eventually identify distance categories implying distinct trip time distributions. However, if so, a reasonable trade-off has to be found between the level of detail considered and the need to be still able to (practically) handle the number of parameters for calibration.

Analogous to the previous section, we now illustrate the comparison of trip distributions towards secondary locations by EOD and the improved model, thus visualize the values shown in Table 6.11 and Table 6.12 with the following Figure 6.7.

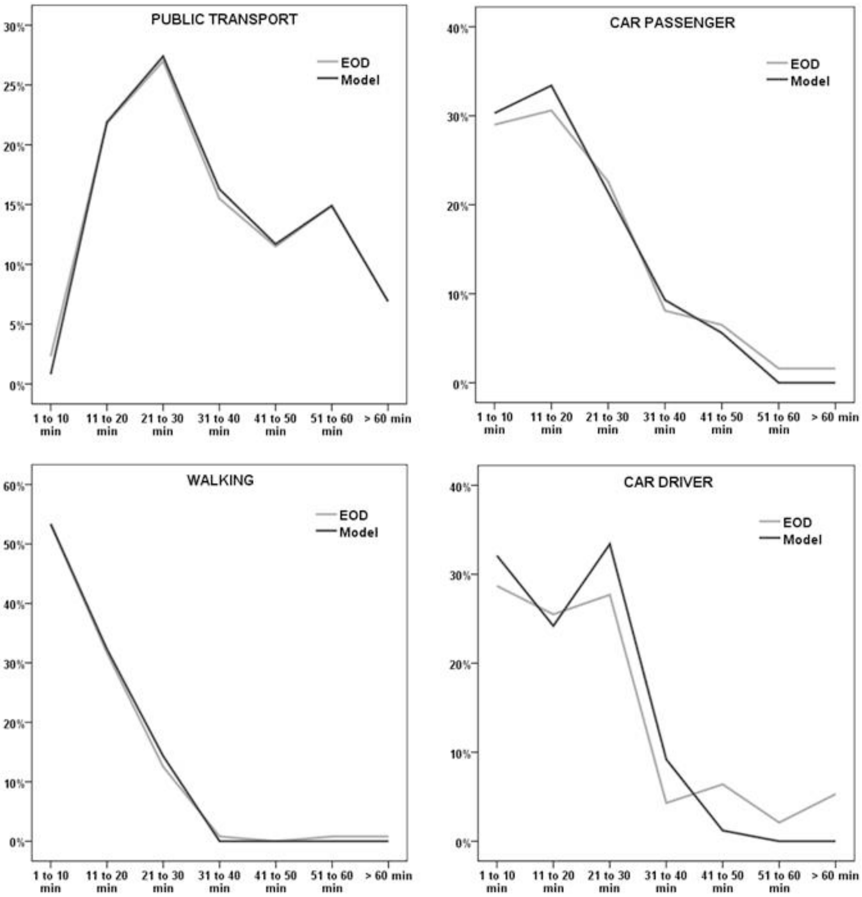


Figure 6.7: Secondary location choice, comparison of EOD and model-2 by mode

The application of the LWPs allows obtaining an almost perfect match between EOD and model for all modes save trips towards secondary activities realized as car driver. This is in particular due to that nearly no alternatives compose the final choice set in more than 40 minutes of travel time distance, despite the adjustment of the underlying travel time matrix. The reason to that could be that these large trips in duration are result of some particular events such as very high congestion levels. Once some trips longer in duration build part of the final choice set, at the same time this will lower the current peak of trips in 21 to 30 minutes travel time distance, resulting in an overall improved match.

Basing on the results shown in Table 6.12 and Figure 6.7, in a next step, we visualize entire spatial paths probabilities to then discuss briefly to what extent we were able to reduce and control combinatorial complexity.

6.4.2 Visualization of spatial path flow

In a final step we now – based on the calibrated destination choice – visualize what is called the spatial path flow. Notice that we use the results obtained of the scenario model run considering adjusted travel times. The following Figure 6.8 depicts one example.

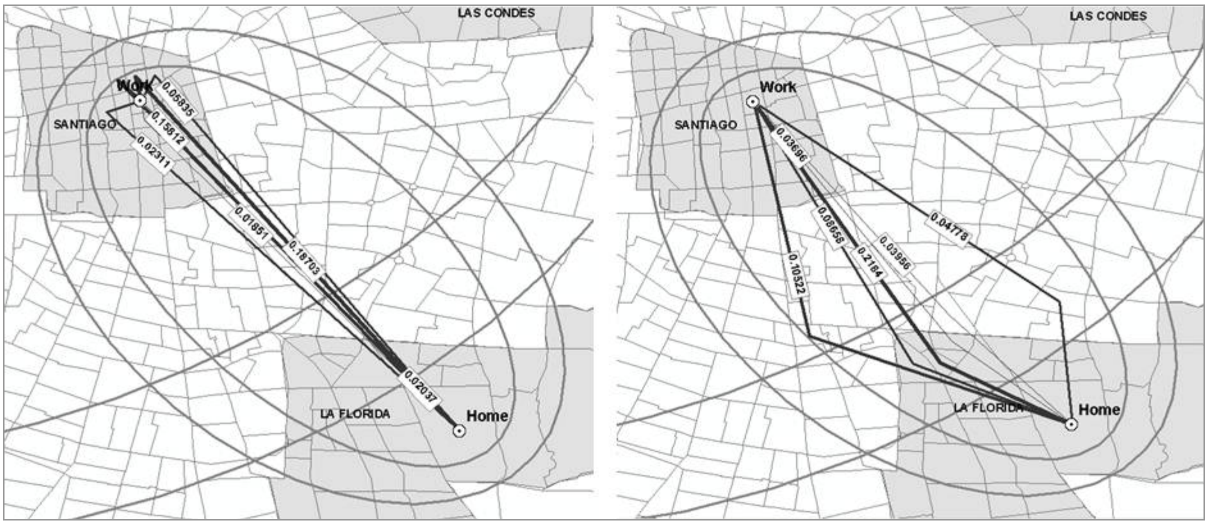


Figure 6.8: Visualization of spatial path flows

Annotation: TpAaTp, TP = 2, UG = 5, left: work-related, right: home-related (TP = Time Period, UG = User Group)

The lines reflect the straight-line connections between home, work and the secondary locations represented by the angle points. Notice that numbers on lines show the probability that a spatial path passes by the secondary location. Respectively, we can summarize the information derived from Figure 6.8 as follows:

- Spatial paths realized by user group 5 (households with no car and a monthly income of 300,000 to 600,000 CHP); paths are either home-related (right) or work-related (left); line labels represent the probability that one out of the alternatives for secondary activities 'Y' is chosen; the probabilities for the home-related paths sum up to 0.53 (or 53%), those for the work-related paths sum up to 0.47 (or 47%); now, the probability of traveling from home to work by public transport can be multiplied with the indicated probabilities.

Notice that we splitted the figure into home- and work-related paths to assure visibility. In fact for every home-work relation spatial paths for both tour relation types are estimated simultaneously. Consider also, that this visualization is theoretically feasible for all OD-relations of the entire city and more information is attached to each path than shown in the figure. In particular, this concerns the attributes introduced in Table 6.5, i.e., the starting times and duration of the activities, the modes used as well as the travel times needed for each individual trip and the complete spatial path. Leaving the example behind and focusing once again on results for the entire city, it is now possible to compare time regimes between the EOD and the model as well as to quantify to what extent the reduction of choice combinations was possible.

6.5 Time regimes: EOD and model

Given the information attached to any spatial path flow, now manifold options for analysis and comparison between EOD and model become feasible. However, we focus on some selected aspects that allow testing the general matching level between model and observations. We begin with the starting times and durations of the primary activity work. The following Figure 6.9 depicts the results comparing model and EOD.

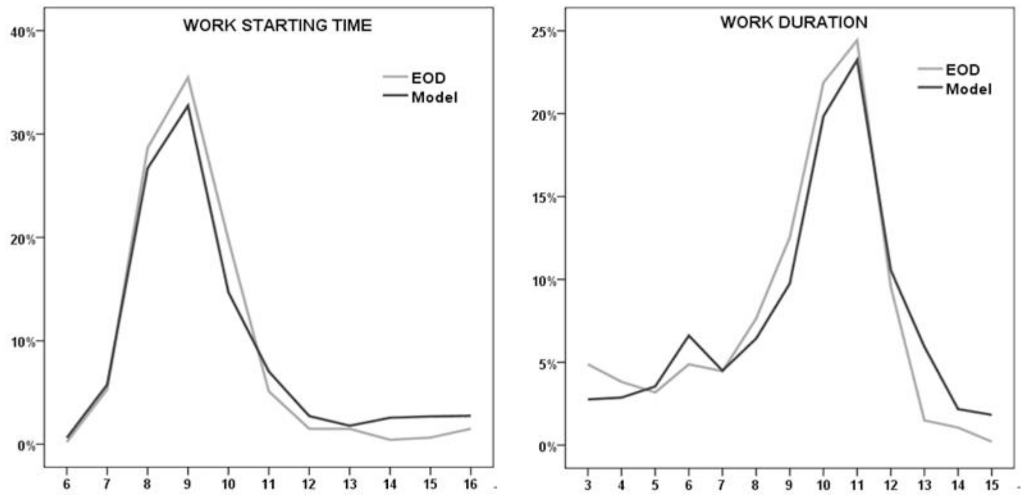


Figure 6.9: Work starting times and durations – EOD and model

Annotation: EOD results are based on pattern type ‘HXYH’ with ‘X’ = work activity. / (N=471)

We observe that in general the model reproduces the observed pattern in the EOD. Essentially, this result confirms that starting times and duration were assigned appropriately. Remember, that we assigned time regimes in accordance to empirical observations and dependent on the user groups making use of a weighted random process (see 3.2).

Another indicator that allows testing the model's ability to reproduce EOD observations is the daily travel time. We applied this indicator in form of a time constraint when it came to the reduction of choice options. On purpose we introduced thresholds for the maximum time spent traveling in accordance to mode combinations. In consequence, one expects that on average the predicted daily travel times by the model would be below observed times, as tours with very long daily travel times were excluded. To cope with this bias comparing model and EOD, in the following Figure 6.10 we also applied the time-space constraints to the EOD data reducing the total number of observations. This explains why the respective N with 380 observations is below the N of the previous figures, although now we consider both pattern types of ‘HXYH’ and ‘HYXH’ for analysis.

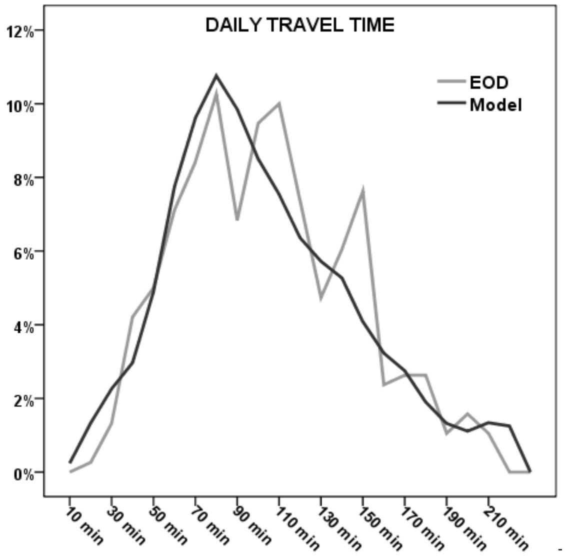


Figure 6.10: Daily travel time: EOD and model

Annotation: EOD results are based on pattern types 'HXYH' and 'HYXH' with 'X' = work activity. In addition, time-space constraints active for the model (daily travel time thresholds and search spaces) were applied to the EOD data, too. / (N=380)

The model is able to reproduce the peak concentrated in 70 to 80 minutes of daily travel time. Mean daily travel time is 100 minutes for the EOD and 97 for the model. The major difference is that the model estimates a continuous decrease until the maximum daily travel time threshold of 215 minutes reflected by mode combination TpTpTp is reached. In contrast, the EOD observations show a set of subsequent peaks, but generally following the same decreasing tendency. This peak-like behavior most probably can be explained once again by the tendency of respondents to round trip times. In general, we can conclude that we obtain a reasonable match between observed and predicted daily travel times. This observation supports also the rules applied for the definition of the final choice set of secondary activities. Especially the definition of the choice set in accordance to empirical trip distributions by mode and travel time ranges assures that also alternatives in further distance are selected resulting in a good reproduction of the observed daily travel times applying the model (see Annex-5 for daily travel time comparisons between model and EOD for further mode combinations).

6.6 Reduction of choice options: quantification

In this final section, we quantify the overall reduction potential of the approach. Remember, that one objective was to reduce the number of feasible options and to control combinatorial complexity. The sequential application of first the time-space constraints, and then the definition of the final choice set made the probabilistic calculation of tour type 'HWYH' possible for the entire city. However, we now can briefly summarize to what extent each methodological step contributed to the reduction of choice options. Notice also that we exclude the temporal dimension at this point as the set of starting times and duration of activities is calculated and assigned once to each spatial path, thus does not increase the number of choice options.

Santiago is split into 618 TAZ. We distinguished between 15 mode combinations and 13 user groups. Respectively, the theoretically feasible maximum amount of combinations for a tour of type 'HWYH' is:

- **618 (H) * 618 (W) * 618 (Y) * 13 (UG) * 15 (MC) = 46,025,661,240 (Level 0).**

Given the home location H, there exist 618 options of work locations (W), another 618 zones to conduct the secondary activity Y leaving from W, divided by 13 user groups (UG) and 15 mode combinations (MC).

However, ESTRAUS commuting matrices that were used as an input to the analysis indicate only 29,305 actual home-work relations, instead of 618^2 = 381,924. Thus, our starting point of the maximum amount of combinations is:

- **29,305 (HW) * 618 (Y) * 13 (UG) * 15 (MC) = 3,531,545,550 (Level 1).**

The first reduction is now implied by the spatial constraint, i.e. by search spaces. In this calculation example we consider search spaces based on larger detour factors of short work durations (see Table 5.8). Thus, on average 94 TAZ constitute the search space (see Table 6.3). This means instead of 618 possible activity sites for 'Y', we can consider on average 94 TAZ resulting in the following number of combinations:

- **29,305 (HW) * 94 (Y) * 13 (UG) * 15 (MC) = 537,160,650 (Level 2).**

The next reduction concerns the daily travel time thresholds. Notice that after the introduction of the spatial constraint each TAZ is calculated separately. Therefore, the reduction effect based on the time constraint was estimated on a sample of 10 zones because the simultaneous calculation becomes difficult considering more zones. This sample indicated that the number of combinations was reduced in result to the time constraint by 60%, leading to:

- **537,160,650 * 0.4 = 214,864,260 combinations (Level 3).**

The final choice set of maximum seven alternatives per primary location further reduces the overall number of combinations as follows. Notice that after 'level 3' all TAZ are once again stored within one single file for the purpose of calibrating the destination choice.

- **12,036,683 (Level 4).**

The final number of about 12 million combinations enters the process of calibration as described before. According to the reductions from level 1 to level 4 we can now quantify the effect of each step. We define level 1 as starting point, i.e. representing 100% of the possible combinations. Next, the spatial constraint (level 2) reduces the combinations to 15.2% of the original size. The application of the temporal constraint (level 3) then is responsible for a further reduction to 6.1%. Finally, the selection of the final choice set in level 4 makes that only 0.3% of the initial amount of choice options remain.

If we transfer this into reduction factors we conclude that the spatial constraint reduces combinatorial complexity by the factor of **6.6**, the temporal constraint by an additional factor of **2.5** and the definition of the final choice set by another factor of **20.3**.

7 Discussion and outlook

The presented model followed a methodology for demand generation on the basis of a fixed activity sequence, assigning locations, modes and times for both activities and trips. The key argument that choices are taken in a hierarchical macro and micro time-space constrained environment underlined both the analysis and the application. The hierarchical choice-making process was exemplified with the concept of macro and micro levels, represented by a tour of type 'HWYH'. Given the multiplicative property of the approach, it was important to prove that the probabilistic estimation of the tour is feasible as soon as time-space constraints are applied to reduce choice options. As a method to control the number of options, so-called search spaces were built based on empirically derived detour factors. These proved to be realistic means for capturing most of the observed trip destinations. However, the estimation of 'HWYH' path flows revealed some discrepancies between the observed trip distributions and those estimated with the model. This disagreement was difficult to overcome given the mismatch between data and model sources. Particularly the car-based travel times appeared to be underestimated by the network model ESTRAUS, thus hampering the attempt to 'match' EOD and model distributions. In consequence, we adjusted travel times and obtained improved results, i.e. a better match for this scenario.

Regarding the research questions raised in one of the opening chapters, the challenge throughout the thesis was twofold: first, the statistical analysis and the attempt to identify the empirical thresholds for the constraints considered; second, 'making them work' within the modeling framework. Without repeating the extensive results provided in the previous chapters, we can briefly resume some main findings that directly affect the calculations of spatial paths:

- The principal time-space constraints identified and included in the approach were the total daily travel time by mode combination and the search space, i.e. the geographical area where a secondary activity is most likely realized. We first applied the spatial constraint to limit the geographical area, while the temporal constraint further reduced the number of options within these boundaries.
- Indeed, the assumption was confirmed that the time spent on the primary activity affects the time-space choice set. Longer duration stays at work limit the search space for conducting additional activities. In addition, we found that the size of the search space and thus the accessible choice set depends on the distance between the primary locations.
- We were able to include a temporal component into the demand generation process. We established a 'chain' of causal relationships, where the search space depended on the time spent at work, while the activity duration defined the expected starting time.
- Considering the constraints applied, the final choice set of at maximum seven options per primary location appears to be a geographically (more) realistic approximation of what location decisions in real life look like. It was possible to consider a set of important criteria when defining the choice sets, which are tour relation, attractiveness and

accessibility of zones as well as controlling that the numbers of alternatives reflect observed trip distributions by mode and travel time.

In its current version the approach is based on what was called before ‘core’ behavioral pattern. On purpose we excluded at different points in this study rarely occurring events from the data when integrating the model parameters. Remember that we defined 15 mode combinations, aggregated time periods of starting times and duration or estimated search spaces based on percentile 70-values for the detour factors. In consequence the approach currently does not reproduce behavior beyond the ‘core’ behavioral pattern we identified through the empirical analysis. This was an important strategy as we were aiming to overcome, especially in the case of destination choice, the problem of having remote choice options with low probabilities. In fact, this strategy permitted to reduce the choice sets substantially, making a probabilistic estimation of tour-based travel demand operational. However, further research is required to identify and at best quantify empirical thresholds between ‘core’ and ‘randomly’ occuring behavioral pattern. This becomes an issue as soon as the presented approach is expanded to other tours or activity pattern. One may implement a model-mechanism that allows coping with these additional events as well by, for instance, including a reduced number of choice options beyond the empirically defined search spaces.

The approach – given its empirical nature – is only partially suitable to predict travel behavioral choices. For instance, we made use of fixed activity pattern and mode combination probabilities. One would expect that changes in the supply quality, i.e. the level-of-services, have an effect on the mode choice probabilities. In this sense it seems beneficial to introduce more dynamic models of travel behavior prediction, for instance, via the use of a multinomial logit model for the mode-destination choice. Nonetheless, already in its current version the approach is able to adjust to several changes in framework conditions. Changes in the input data will lead to alterations in the distribution of spatial path flows. This concerns changes of land-use (predicted by land-use model MUSSA), the home-work related decisions for locations and modes (predicted by transport model ESTRAUS) or the transport supply (again ESTRAUS). In addition, changes in work time regimes will lead to different search spaces, and thus to different choice sets for secondary activities locations. Travel time savings due to expansion of infrastructure or operational adjustments in public transport will lead to more secondary activities reachable within search spaces.

Not surprisingly, we can conclude that the approach in its current version offers possibilities for further improvement and analysis. Aspects addressing potential future research and dealing with strengths and weaknesses of the approach are subject of the following discussion along some selected issues perceived as of major relevance.

7.1 Transferability of the approach

Several issues come up when thinking about the potential of transferability of the approach. Positive characteristic is that commonly available data was used. Essentially, we assumed availability of commuting matrices and information provided by a traditional travel survey.

We suppose that such information is on-hand in many other cities. In addition, the methodology suggested for the definition of a reduced choice set is generic by nature and thus is transferable to another context, i.e. another city.

A bigger challenge is to transfer the approach to further activity pattern and tours. Though in this thesis we focused on one specific tour, there are many ‘lessons learnt’ that allow for an extension of the methodology to other patterns. Generally, two major issues come up in this sense: first, the calculation of a different single tour, say of type ‘HYH’ or ‘HXYYH’. These examples already are quite different regarding their modeling complexity. A tour of type ‘HYH’ implies that no interrelations between activities need to be addressed. In line with the approach pursued in this thesis, the task would be to define a search space but only based on one anchor point, the home location. Similar questions would come up, such as whether the extension of this search space depends on the starting and/or activity time spent at the location or on the transport modes used. In case of a tour of type ‘HXYYH’, an additional challenge pops up: the modeling of two secondary activities in a row, namely, the time-spatial relation of ‘YY’. However, the concept of search spaces might be applicable in these cases as well. The subject of further investigation would be to find out whether the search space alters due to the additional activity or if the second ‘Y’ is most likely realized within the same geographical boundaries as for ‘HXYH’ because the time constraint applies as well. In addition, one will have to define whether the second activity ‘Y’ most likely occurs in the proximity to home location ‘H’ or the previous activity ‘Y’. If we then assume that two choice sets for both activities ‘Y’ (say, seven alternatives each) are defined, a larger number of combinations for spatially related ‘Ys’ has to be processed. Applying the constraints considered in this thesis, an evaluation of these combinations could be done regarding their compliance with the total daily travel time restrictions. In addition, it is a new challenge to consider different mode combinations with more activities conducted. The related research question would then be to what extent further differentiated mode combinations are still required or whether in the case of ‘HXYYH’ it is reasonable to assume that identical modes are used for ‘XY’, ‘YY’ and ‘YH’.

Another concern is the consideration of inter-tour relationships, which means the task to consider the decisions taken or resources spent at the first tour for the definition of the constraints for the second or third tours. So far, inter-tour relationship is a strength of activity-based simulations while trip-chain or tour-based models typically do not cover this aspect. Regarding the methodology proposed here, a feasible approach would be to limit time (for activities and trips) for a second tour according to the time spent on a previous tour and thus to reduce the spatial coverage of the search space. Of course, additional analytical efforts are required to define empirical evidence for the suggestions made in this and in the previous paragraph. However, if we assume an expansion of the current approach to single tours with maximum two ‘Y’ activities in a row and a maximum of four consecutive tours per day, over 90% of the observed activity pattern variations in Santiago would become reproducible.

7.2 The magical number seven

The final choice set was limited to maximum seven alternatives per primary location. Adjacent disciplines like psychology made an important contribution to the idea of 'cognitive restriction' introduced in the opening chapters. However, if we transfer this to transport demand analysis, subsequent questions arise. Is it universally correct to assume seven alternatives per primary location? When is a primary location sufficiently known to the individual so that we can reasonably assume that seven choice options for secondary activities are likely to be identified by the individual? What is the minimum distance between two primary locations that allow us to consider their surrounding different from each other? The latter question addresses the situation where, for instance, home and work are in short distance from each other. From a behavioral point of view, there is no reason why an individual would evaluate two choice option sets according to the two locations of home and work. It is rather likely that in this case one search space containing both primary locations has to be considered. This brings up the question whether still seven (or more or less) alternatives would be considered. The much the magical number seven intuitively feels 'right', the notion remains that the number of a final choice set also depends on the distance between primary locations.

7.3 Suggestions for further development

Travel Times: Several aspects covering the current methodology as well as data and information used can be improved. Doubts remain how reliable are the travel times provided by Santiago network model ESTRAUS. The identified discrepancies especially for car-based travel times between EOD and ESTRAUS suggest revising times in more detail. Latest with the next EOD travel survey, this comparison between observed and modeled travel times seems recommendable.

Locations capacities: In the current version of the model we did not consider the capacity constraints of locations. After the trip distribution it might come up that certain locations attract more trips than their capacity allows. To consider this in the application iterations are required. After the first trip distribution it has to be evaluated to what extent TAZ are 'saturated' and a shift of trip probabilities to other alternatives of the final choice set becomes necessary. However, this issue raises a new set of research questions that could not be handled with in this thesis. For instance, how much do the attractiveness of shopping facilities drop with 60, 70 or 80% of saturation? In consequence the subsequent question would be how much a neighboring, not-saturated TAZ, can benefit from the decreasing attractiveness of its neighbor zone. In addition, solutions to this problem depend very much on the spatial context. The more aggregated the perspective, the less unclear it is how effects of altering attractiveness actually occur. While we still might quantify how much a cinema benefits from its crowded neighbor, this gets difficult – honestly we might say not possible – if we try for instance to quantify to what extent smaller retail stores benefit from an extremely crowded shopping mall. These examples already indicate that much more empirical research in this area and its use within models of transport demand seems relevant and needed.

Iterate supply-demand: The issue of iterations brings up another concern, which has not been addressed yet: the update of LOS matrices after trip assignment and the feedback of actualized travel times into the steps of demand generation. This was not possible as only a fraction of the demand, i.e. a tour of type 'HWYH', has been modeled so far and the access to Santiago network model has been limited. However, once we assume an entire set of observed activity pattern is modeled and thus the complete trips demand is assigned to the networks, iterations between demand and supply become feasible. Moreover, the approach in its current design includes starting times and duration of activities, which allows to feedback time-of-day dependent LOS. Nonetheless, more research is required to actually transfer the current methodology to further tour types (see discussion above) and test the possibilities of bringing demand and supply in a state of equilibrium.

Calibration: With the expansion of the methodology to other tour types, another issue comes up dealing with the calibration of destination choice. In case of one macro-micro activities combination, we were able to assume the macro decisions (home-work relation) as fixed and needed to adjust only the secondary activity location choice. This appears more challenging if two secondary activities in a row ('YY') occur. As long as each 'Y' is related to another spatially fixed primary activity ('H' or 'W'), it is possible to adjust destination choice similarly to the way described in the previous chapter. If this spatial relation is not given, it seems recommendable to consider also tour-based indicators for calibration rather than the trip-based. An example could be to reproduce the total daily travel time of a tour rather than the single trip travel time distributions.

Random behavior: One of the underlying objectives of the statistical analysis conducted was to identify what we defined as 'core behavioral patterns'. The attempt was to capture the majority of behavioral heterogeneity at the cost of excluding 'randomly' occurring behavior, for instance, a 5-hour daily travel time, a 10-hour work activity starting at 20h in the evening or a 120-minute walking trip to a secondary location. As expected, it was possible to identify these behavioral 'cores'. However, the definition of a clear cut between 'core' and 'random' behavior through empirical thresholds makes the model currently insensitive for seldomly occurring events. At the moment it is not foreseen that a person (or user group) behaves 'randomly', e.g., a walking trip lasting 120 minutes is not a part of feasible choice options. Partly due to this exclusion it was difficult to match EOD and model distributions when it came to destination choice calibration. Hence, one could attempt to integrate a limited number of one or two randomly selected alternatives into the final choice set not complying with empirical thresholds.

Search spaces: Another option for improvement concerns the estimation of search spaces. Currently, they are based on the straight-line distances between the home and work locations. Obviously, this is only a rough approximation of what traveled routes are in reality. The validation regarding matching levels between EOD and search spaces revealed that quite clearly: matching levels descended in the order from walking trips (likely similar to straight-line distance), to car-travel and, finally, to public transport where trips on the networks are less likely to be close to straight-line connections. Further analysis is required to clarify whether the non-matching cases tend to be due to ‘random behavior’ or are caused by the search spaces that not follow networks of private and public transport. Intuitively, it appears rational to consider routes chosen on networks for the estimation of search spaces, making location choice even more behaviorally realistic. However, this attempt is not free of potential bias. As long as detailed time-space trajectories of individuals are not available in greater amounts, there remains only the option of routing OD relations using network models, for instance applying algorithms of identifying shortest (time or space) path distances. It was subject of discussion that these techniques still provide only an approximation of what route choices are in real life. Today new technologies, e.g., smartphone applications, are available to measure and to map individual trajectories. However, important issues remain to be solved. This refers to data security concerns as individual behavior becomes traceable. Practical challenges remain with tracked trajectories when starting and ending locations of specific activities and trips have to be defined. ‘Passive tracking’ of trajectories (without surveying individuals) does not yet offer the information quality necessary for estimating behaviorally more realistic search spaces.

Zoning system: There was an interrelation identified between the geographical zoning system and the obtained results. Remember that in one case a large TAZ of the home location made that no choice alternatives were found in 10 minutes walking distance, because the intrazonal travel time was estimated to be longer than 10 minutes. To overcome this, one could consider different zone sizes depending on the realized activity. For instance, macro scale locations (home, work) can be searched at a macro TAZ level, while micro scale locations (shopping, leisure) can be searched at a micro level, i.e. at a level of residential blocks. However, this increases the number of possible locations within search spaces at the cost of search efficiency, as more alternatives have to be evaluated (weighted, ranked). Nonetheless, this appears an interesting option as soon as search spaces are better refined, i.e. spatially reduced. The methodology by design can easily be extended to a bi-level of geographies, say to a macro-micro spatial hierarchy.

User groups: Another relevant issue is the segmentation of the demand in household user groups. These groups had a strong influence on the model structure and on the data analysis. Spatial path flows were estimated for user groups and primary activity duration was found to be statistically correlated with them. Originally, their adoption was motivated by the fact that OD trip flows by mode for the home-work relation were available from ESTRAUS and population segmented by user groups from land-use model MUSSA. Nonetheless, it should be noticed that the methodology does not depend on this segmentation. Once different data sources are available for the reproduction of commuting relations, it becomes possible to vary population segmentations as well.

Analytical modeling: So far, the model has been based on empirical probabilities calculated directly from data. In a further developed version, respective distributions may be derived from functional probabilities. This applies, for instance, to the calibration of the destination choice model where a set of weighting parameters is currently applied to reproduce observed distributions. In addition, discrete-choice models may be applied for the calibration of parameters of the mode and destination choice basing on the identified subset of alternatives. Currently, we already apply a random process to introduce some choice variations and by that simulate that the attractiveness of alternatives does not solely depend on their land-use characteristics. Advanced choice models can bring in more behavioral realism into the evaluation of alternatives, considering additional socio-economic attributes of individuals (or user groups) as well.

Computational issues: Eventually, a potential remains for applying a (technically) more efficient manner of calculating the spatial path flows. As said before, data analysis as well as model development, application and testing were realized using SPSS software. This was convenient given that the software supports programming by providing pre-prepared, easy-to-learn syntax commands. However, the program is not explicitly designed for the application of complex and iterative transport demand modeling. Nonetheless, a positive feature of the current calculation structure is that the model works in a sequential manner, calculating stepwise the TAZ for entire Santiago. This stepwise approach allows calculating TAZ in parallel by using different computers at a time and thus making the calculation faster.

7.4 Closing remarks

In one of the opening chapters it was argued that the model represents a sort of ‘hybrid’ approach considering aspects of different model types. We can conclude that the model provides clearly more information than traditional four-step-models but less than activity-based approaches. This position ‘in between’ gets evident when looking at some of the model’s major characteristics: the approach by design offers the option to iterate between demand and supply, thus reaching a unique solution (equilibrium); it is based on aggregated demand user groups but considers the interdependencies of activity and travel related choices; it considers time-space constraints resulting in a more realistic reproduction of travel choices. The concept of search spaces as result of the daily constraint behavior is applicable both for aggregated and disaggregated models of transport demand. We used ellipses to reduce choice options and complexity at the same time increasing behavioral realism. Disaggregated models including agent-based microsimulators can make use of this concept as well to reduce choice complexity and define a most feasible set of choice alternatives.

With regard to our example of tour type ‘HWYH’, we can derive some contributions of general validity. Aspects of the hierarchical choice process for activities, locations and modes are applicable in activity-based models as well. Again, one may adopt the concept of estimating search spaces and apply it as first step of the destination choice model. In the context of activity-based approaches where individual decision-making is modeled, the search spaces might become much more individualized considering more attributes (socio-economic, supply-side) during their definition. It also appears straightforward to consider maximum thresholds of mode-dependent tour travel times within a micro simulation.

Eventually, through the final choice set the approach provides a strongly reduced number of potential destination choices that at the same time comply with the imposed constraints. This subset can serve as a set of alternatives taken into account when applying discrete-choice models of mode and destination choice. Often these subsets of alternatives are taken randomly from the entire set of options available, i.e. from the entire study area. The model we developed for the choice of secondary activity ‘Y’ is then a much more precise manner of selection of this subset of alternatives.

References

(1) Axhausen, K., 2008. Definition of movement and activity for transport modelling. In: D. A. Hensher and K.J. Button, Eds. *Handbook of Transport Modelling.* 2nd ed. Amsterdam: Elsevier, pp. 329-343.

(2) Bates, J., 2008. History of demand modelling. In: D. A. Hensher and K. J. Button, Eds. *Handbook of Transport Modelling*. 2nd ed. Amsterdam: Elsevier, pp. 11-34.

(3) Bhat, C.R. and Pinjari, A.R., 2008. Duration modeling. D. A. Hensher and K.J. Button, Eds. *Handbook of Transport Modelling*. 2nd ed. Amsterdam: Elsevier, pp. 105-131.

(4) Bowman, J.L., 1998. *The day activity schedule approach to travel demand analysis. PhD Dissertation*. Massachusetts Institute of Technology.

(5) Bowman, J.L. and Ben-Akiva, M.E., 2000. Activity-Based disaggregate travel demand model system with activity schedules. *Transportation Research Part A*, 35*(1)*, pp. 1-28.

(6) Bowman, J.L. and Bradley, M., 2008. Activity-based models: approaches used to achieve integration among trips and tours throughout the day. In: *European Transport Conference*. Leeuwenhorst, Netherlands, October 2008.

(7) Bradley, M. and Vovsha, P., 2005. A model for joint choice of daily activity pattern types of household members. *Transportation*, 32, pp. 545-571.

(8) Bradley, M., Bowman, J.L. and Griesenbeck, B., 2010. SACSIM: an applied activity-based model system with fine-level spatial and temporal solution. *Journal of Choice Modelling*, 3*(1)*, pp. 5-31.

(9) Brosius, F., 2008. *SPSS-Programmierung. Effizientes Datenmanagement und Automatisierung mit SPSS-Syntax*. 2n ed. Heidelberg. mitp.

(10) Buliung, R.N., 2005. Activity/Travel Behaviour Research: approaches and Findings with Identification of Research Themes and Emerging Methods. Centre for Spatial Analysis, McMaster University *(Working Paper Series)*, unpublished.

(11) Buliung, R.N. and Kanaroglou, P.S., 2006a. A GIS toolkit for exploring geographies of household activity/travel behavior. *Journal of Transport Geography*, 14*(1)*, pp. 35-51.

(12) Buliung, R.N. and Kanaroglou, P.S., 2006b. Urban Form and Household Activity-Travel Behaviour. *Growth and Change*, 37*(2)*, pp. 172-199.

(13) Buliung, R.N. and Kanaroglou, P.S., 2007. Activity-Travel Behaviour Research: conceptual Issues, State of the Art, and Emerging Perspectives on Behavioural Analysis and Simulation Modelling. *Transport Reviews*, 27*(2)*, pp. 151–187.

(14) Cascetta, E., 2009. *Transportation systems analysis: models and applications*. 2nd ed. New York. Springer.

(15) Caussade, S., Ortúzar, J. de D., Rizzi, L. and Hensher, D., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B*, 39, pp. 621-640.

(16) Cea, J. De, Fernández, E., Dekock, V., Soto, A. and Friesz, T., 2003. ESTRAUS: a computer package for solving supply-demand equilibrium problems on multimodal urban

transportation networks with multiple user classes. In: *82nd Transportation Research Board Annual Meeting*. Washington D.C., 12-16 January 2003.

(17) Cirillo, C., Cornelis, E., Legrain, L. and Toint, P.L., 2003. Combining spatial and temporal dimensions in destination choice models. In: *European Transport Conference*, Strasbourg, 8-10 October 2003.

(18) Citilabs, 2010. Land use forecasting. [online] Available at: <www.citilabs.com/cube-land.html> [Accessed 25.04.2011]

(19) Davidson, W., Donnelly, R., Vovsha, P., Freedman, J., Ruegg, S., Hicks, J., Castiglione, J. and Picado, R., 2007. Synthesis of first practices and operational research approaches in activity-based travel demand modeling. *Transportation Research Part A*, 41*(5)*, pp. 464-488.

(20) DICTUC – Dirección de Investigaciones Científicas y Tecnológicas de la Pontificia Universidad Católica de Chile, 2001. *Encuesta Origen Destino 2001: Informe Final*. Santiago de Chile.

(21) Dijst, M. and Vidakovic, V., 1997. Individual action space in the city. In: D.F. Ettema and H.J.P. Timmermans, Eds. *Activity-based approaches to travel analysis*. Oxford: Elsevier, pp. 117-134.

(22) Doherty, S.T., 2006. Should we abandon activity type analysis? Redefining activities by their salient attributes. *Transportation*, 33, pp. 517-536.

(23) Doherty, S.T. and Mohammadian, A., 2011. The validity of using activity type to structure tour-based scheduling models. *Transportation*, 38, pp. 45-63.

(24) Dugge, B., 2006. Ein simultanes Erzeugungs-, Verteilungs-, Aufteilungs- und Routenwahlmodell (EVA-U). *Schriftenreihe des Instituts für Verkehrsplanung und Straßenverkehr*, 9.

(25) Ettema, D., Bastin, F., Polak, J. and Ashiru, O., 2007. Modelling the joint choice of activity timing and duration. *Transportation Research Part A*, 41, pp. 827-841.

(26) Fan, Y. and Khattak, A., 2008. Urban form, individual spatial footprints, and travel: examination of space-use behaviour. *Transportation Research Record*, *2082*, pp. 98-106.

(27) Fellendorf, M., Haupt, T., Heidl, U. and Scherr, W., 1997. Activity based demand forecasting in daily practice. In: D.F. Ettema and H.J.P. Timmermans, Eds. *Activity-based approaches to travel.* Oxford: Elsevier, pp. 55-72.

(28) Florian, M., Wu, J. H. and He, S., 2002. A multi-class multi-mode variable demand network equilibrium model with hierarchical logit structures. In: M. Gendreau and P. Marcotte, Eds. *Transportation and Network Analysis: Current trends*. Dordrecht: Kluwer Academic Publishers, pp. 119-133.

(29) Frank, L., Bradley, M., Kavage, S., Chapman, J. and Lawton, T.K., 2008. Urban form, travel time, and cost relationship with tour complexity and mode choice. *Transportation*, 35, pp. 37-54.

(30) Friedrich, M. and Schiller, C., 2010. Modellierung von Verkehrsangebot und Verkehrsnachfrage. *Weiterbildungskurs*. Technische Universität Dresden, unpublished.

(31) Golob, T.F. and Hensher, D.A., 2007. The trip chaining activity of Sydney residents: a cross-section assessment by age group with a focus on seniors. *Journal of Transport Geography*, 15*(4)*, pp. 298-312.

(32) Hägerstrand, T., 1970. What about people in Regional Science? *Papers in Regional Science*, 24 *(1)*, pp. 6-21.

(33) Hensher, D.A. and Button, K.J., 2008. *Handbook of Transport Modelling*. 2nd ed. Amsterdam. Elsevier.

(34) Hensher, D. and Mannering, F., 1994. Hazard-based duration models and their application to transport analysis. *Transport Reviews*, 14*(1)*, pp. 63-82.

(35) Iacono, M., Levinson, D. and El-Geneidy, A., 2008. Models of Transportation and Land Use Change: a Guide to the Territory. *Journal of Planning Literature*. 22*(4)*, pp. 323-340.

(36) Justen, A., Martínez, F. and Cortés, C., 2010. Time and spatial dependent activities and travel choices: results from Santiago de Chile. In: *XVI Panamerican Conference of Traffic and Transportation Engineering and Logistics (PANAM)*. Lisbon, 15-18 July 2010.

(37) Kim, H-M. and Kwan, M-P., 2003. Space-time accessibility measures: a geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration. *Journal of Geographical Systems*, *5*, pp. 71-91.

(38) Kitamura, R., 1997. Applications of models of activity behavior for activity based demand forecasting. [online] Arlington TX: Texas Transportation Institute. Available at: <tmip.fhwa.dot.gov/resources/clearinghouse/docs/abtf/kitamura.pdf>. [Accessed 25.04.2011]

(39) Lohse, D., 1997. Ermittlung von Verkehrsströmen mit n-linearen Gleichungssystemen unter Beachtung von Nebenbedingungen einschließlich Parameterschätzung. Verkehrsnachfragemodellierung: Erzeugung, Verteilung, Aufteilung. *Schriftenreihe des Instituts für Verkehrsplanung und Straßenverkehr*, *5*.

(40) Malhotra, N., 1982. Information load and consumer decision making. *The Journal of Consumer Research*, *8*, pp. 419-430.

(41) Martínez, F.J., 1996. MUSSA – land use model for Santiago City. *Transportation Research Record*, *1552*, pp. 126-134.

(42) Martínez, F., 2008. Towards a land use and transport interaction framework. In: D. A. Hensher and K.J. Button, Eds. *Handbook of Transport Modelling*. 2nd ed. Amsterdam: Elsevier, pp. 181-201.

(43) Martínez, F., Justen, A. and Cortés, C., 2009a. A bi level model of daily trip chains. In: *International Time Use Observatory Workshop (TUOW)*. Santiago de Chile, Chile.

(44) Martínez, F., Justen, A. and Cortés, C., 2009b. Methodology for an integrated modelling of macro and microscopic processes in urban transport demand. In: *12th International Conference of the International Association for Travel Behaviour Research.* Jaipur, India, 13-18 December 2009.

(45) McNally, M. and Rindt, C., 2008. The activity-based approach. D. A. Hensher and K.J. Button, Eds. *Handbook of Transport Modelling*. 2nd ed. Amsterdam: Elsevier, pp. 55-73.

(46) Miller, G.A., 1956. The magical number seven, plus minus two: some limits on our capacity for processing information. *Psychological Review*, *101*, pp. 343-352.

(47) Miller, H.J., 2004. Activities in Space and Time. In: *Handbook of Transport 5: Transport Geography and Spatial Systems*. Oxford: Elsevier, pp. 647-658.

(48) Miller E.J., Roorda, M.J. and Carrasco J.A., 2005. A tour-based model of travel mode choice. *Transportation*, *32*, pp. 399-422.

(49) MIDEPLAN – Ministerio de Planificación, 2002. *Estudio análisis políticas de usos de suelo: Informe Final.* Universidad de Chile. Santiago de Chile.

(50) MIDEPLAN – Ministerio de Planificación, 2005. *ESTRAUS: Manual del Usuario, Version 5.5*. Santiago de Chile.

(51) Newsome, T.H., Walcott, W.A. and Smith, P.D., 1998. Urban activity spaces: illustrations and application of a conceptual model for integrating the time and space dimensions. *Transportation*, *25*, pp. 357-377.

(52) Pendyala, R.M., Yamamoto, T. and Kitamura. R., 2002. On the formulation of time-space prisms to model constraints on personal activity-travel engagement. *Transportation*, *29*, pp. 73-94.

(53) Petersen, E. and Vovsha, P., 2006. Intra-household car type choice for different travel needs. [CD-ROM]. In: *85th Transportation Research Board Annual Meeting*. Washington D.C., 22-26 January 2006.

(54) Schiller, C., 2007. Erweiterung der Verkehrsnachfragemodellierung um Aspekte der Raum- und Infrastrukturplanung. *Schriftenreihe des Instituts für Verkehrsplanung und Straßenverkehr*, *10*.

(55) Schönfelder, S. and Axhausen, K.W., 2003. Activity spaces: measures of social exclusion? *Transport Policy*, 10*(4)*, pp. 273-286.

(56) SECTRA – Secretaría de Planificación de Transporte, 2002. *Mobility Survey EOD 2001: Executive Report*. Santiago de Chile.

(57) SECTRA – Secretaría de Planificación de Transporte, 2008. *EFFUS - Efectos urbanos futuros en Santiago: Exploración aplicada de los cambios en el uso de suelo generados por Transantiago en los años 2010 y 2015, estimados con los modelos MUSSA y ESTRAUS*. Santiago de Chile.

(58) Schiller, C., 2007. Erweiterung der Verkehrsnachfragemodellierung um Aspekte der Raum- und Infrastrukturplanung. *Schriftenreihe des Instituts für Verkehrsplanung und Straßenverkehr*, *10*.

(59) Timmermans, H., Arentze, T. and Joh, C-H., 2002. Analysing space-time behaviour: new approaches to old problems. *Progress in Human Geography*, 26*(2)*, pp. 175-190.

(60) TRB Special Report 288, 2007. Metropolitan travel forecasting: current practice and future direction. Transportation Research Board, 2007. Washington, DC.

(61) Vovsha, P., Bradley, M. and Bowman, J.L., 2004. Activity-based travel forecasting models in the United States: Progress since 1995 and prospects for the future. In: *EIRASS Conference on Progress in Activity-Based Analysis*. Vaeshartelt Castle, Maastricht, Netherlands, 28-31 May 2004.

(62) Witlox, F., 2007. Evaluating the reliability of reported distance data in urban travel behaviour analysis. *Journal of Transport Geography*, *15*, pp. 172-183.

(63) Zegras, C., 2004. Influence of Land Use on Travel Behaviour in Santiago, Chile. *Transportation Research Record: Journal of the Transportation Research Board*, *1898*, pp. 175-182.

(64) Zumkeller, D., Manz, W., Last, J. and Chlond, B., 2005. Die intermodale Vernetzung von Personenverkehrsmitteln unter Berücksichtigung der Nutzerbedürfnisse (INVERMO). *Schlussbericht*. Institut für Verkehrswesen - Universität Karlsruhe (TH), Karlsruhe.

Annex

Annex-1: Aggregation of travel purposes

The small number of observations for some activities in the EOD makes it indispensable to develop a strategy for their aggregation. Otherwise too few observations are available for the calculation of empirical distributions of the EOD. A systematic and traceable method to aggregate travel purposes of the EOD is developed in this section. The objective is twofold: to join activities that show clear similarities and to build up groups that represent a substantial number of observations. The aggregation was realized based on the analysis of some selected characteristics of each purpose. The following Annex-Table 1 and Annex-Table 2 provide an overview of the considered characteristics of activity duration, trip duration and modal split.

Annex-Table 1: Temporal characteristics of travel purposes in the EOD

			Activity Duration (in min)			Trip Duration (in min)		
	Number	Valid N in %	Median	Mean	Standard Deviation	Median	Mean	Standard Deviation
Shopping	16,215	21.9	20	42	53	10	15	18
Work	14,461	19.5	530	472	204	33	40	28
Education	9,047	12.2	325	341	125	20	25	21
Visit	7,738	10.4	150	197	165	18	26	27
Other	6,732	9.1	75	122	144	15	23	26
Leisure	5,645	7.6	120	155	126	15	22	30
Bring somebody	5,126	6.9	5	21	56	15	20	28
Services	4,129	5.6	45	72	82	25	30	23
Health	1,887	2.5	78	104	89	25	31	22
Eat	1,310	1.8	57	82	69	10	15	25
Business	1,122	1.5	155	211	187	30	34	38
Bring something	690	0.9	15	41	72	15	19	20

Annex-Table 2: Modal split by travel purposes in the EOD

	Car Driver		Car Passenger		Public Transport		Taxi	
	Valid N	Share in %	Valid N	Share in %	Valid N	Share in %	Valid N	Share in %
Work	2,797	19.3	985	6.8	7,257	50.2	114	0.8
Business	381	34.0	107	9.5	366	32.6	23	2.0
Education	180	2.0	1,226	13.6	3,242	35.8	40	0.4
Shopping	1,597	9.8	1,432	8.8	2,392	14.8	103	0.6
Visit	879	11.4	1,329	17.2	2,061	26.6	108	1.4
Other	343	5.1	1,505	22.4	1,552	23.1	104	1.5
Services	501	12.1	256	6.2	2,062	49.9	54	1.3
Leisure	458	8.1	881	15.6	978	17.3	61	1.1
Health	113	6.0	209	11.1	969	51.4	79	4.2
Bring somebody	1,592	31.1	564	11.0	573	11.2	54	1.1
Eat	198	15.1	188	14.4	140	10.7	22	1.7
Bring something	149	21.6	83	12.0	121	17.5	13	1.9
	Walking		**Shared Taxi**		**Bicycle**		**Other**	
	Valid N	Share in %	Valid N	Share in %	Valid N	Share in %	Valid N	Share in %
Work	2,062	14.3	319	2.2	544	3.8	383	2.6
Business	171	15.2	22	2.0	35	3.1	17	1.5
Education	3,053	33.7	109	1.2	55	0.6	1,142	12.6
Shopping	10,021	61.8	361	2.2	266	1.6	43	0.3
Visit	2,902	37.5	134	1.7	234	3.0	91	1.2
Other	2,884	42.8	218	3.2	51	0.8	75	1.1
Services	899	21.8	238	5.8	81	2.0	38	0.9
Leisure	2,881	51.0	77	1.4	247	4.4	62	1.1
Health	372	19.7	117	6.2	6	0.3	22	1.2
Bring somebody	2,183	42.6	60	1.2	73	1.4	27	0.5
Eat	716	54.7	14	1.1	23	1.8	9	0.7
Bring something	267	38.7	18	2.6	32	4.6	7	1.0

Annotation: The category 'Other' comprises trips realized by a) motorcycle (driver, passenger), b) school transport, c) institutional transport and d) other transport which was not further defined.

To avoid any ad hoc aggregation, an analytic step is introduced to quantify similarity and, respectively, dissimilarity between all activities. The idea is to quantify the distance between all activities to consider a more robust method when it comes to grouping. Otherwise, the grouping would be based on a somehow intuitive approach where, for instance, the purpose 'Eat' is grouped together with the purpose 'Leisure' without any testimony to the assumed likeness. For this purpose, the characteristics of activity duration, trip duration and modal split were considered. Additionally, each aspect is weighted in such a way that differences in activity duration times are more relevant, followed by simply weighted travel times and modal split values. Notice that the use of weighting factors is done in order to reflect that, for instance, the time spent on an activity is assumed to be related to the resources one is willing to invest. It was already argued in section 4.3.2 that activity duration plays a crucial role in the sequence how activities are planned during the day (Doherty and Mohammadian, 2011, p. 57). Respectively, an on average longer activity is associated with longer-term resources investment and thus is considered to be of higher priority. The most palpable example is once again the work activity where a long-term time investment (going to school, to university, etc.) is related to generally high average daily activity durations (see Annex-Table 1).

Concerning the consideration of weights, we multiply activity duration by 2, the trip duration time by 1 and each of the eight transport modes by the factor of 0.125 summing up to 1. Notice that the values of 2, 1 and 0.125 are heuristic and are used to insert a differentiated weighting scheme. This also means that they are free of empirical evidence and are inevitably arbitrary.

The distance D is based upon the median values of activity and trip duration to avoid the influence of statistic outliers. Durations are measured in minutes and the modal split in percentage (see Annex-Table 1 and Annex-Table 2). To define a common unit, the temporal values for activity and trip duration were converted into percentages relating them to the maximum duration of 1,440 minutes representing the 24h-day. The calculation of D is exemplified with the following example of the distance between the work and the shopping activity and is defined as the radical of:

- $D = 2*(530/1440*100 - 20/1440*100) + (33/1440*100 - 10/1440*100) + 0.125*(19.3 - 9.8) + 0.125*(14.3 – 61.8) + 0.125*(6.8 - 8.8) + 0.125*(50.2 - 14.8) + 0.125*(0.8 – 0.6) + 0.125*(2.2 – 2.2) + 0.125*(3.8 - 1.6) + 0.125*(2.6 – 0.3) = 84.8$

This calculation is used to estimate a dissimilarity index for the relations between all EOD travel purposes as shown in the following Annex-Table 3.

Annex-Table 3: Distance matrix between EOD travel purposes

	Work	Business	Education	Shopping	Visit	Other	Services	Leisure	Health	Bring Somebody	Eat	Bring Something
Work	-	57.2	38.4	84.8	62.4	75.9	70.8	69.8	67.9	85.3	79.5	81.3
Business		-	33.5	31.8	9.0	22.6	22.5	16.7	18.8	29.1	26.1	27.5
Education			-	52.6	29.6	41.1	47.0	36.3	41.7	54.6	47.2	51.3
Shopping				-	24.8	14.1	15.2	17.6	20.9	8.4	8.2	6.8
Visit					-	13.7	22.1	8.1	18.5	26.5	18.8	22.6
Other						-	14.3	10.0	9.8	16.4	8.6	13.5
Services							-	21.4	7.0	17.4	13.8	13.8
Leisure								-	17.0	21.7	11.9	18.6
Health									-	23.1	16.2	19.4
Bring Somebody										-	11.7	4.7
Eat											-	10.8
Bring Something												-

As a result a half-filled matrix of distance values is calculated since the distance, for instance, between work and shopping is the same as between shopping and work. The lowest distance is measured between the travel purposes of ‘Bring Somebody’ and ‘Bring Something’ with 4.7 (grey cell), the highest - between ‘Work’ and ‘Bring Somebody’ with 85.3 (black cell). The distance values are now used to support the definition of aggregation rules. With ordering the matrix by columns, a ranking order was established with ascending distance values to identify which activities appear (relatively) similar to each other (not shown here). The first glance at the Annex-Table 3 makes it evident that the highest distance values are to the ‘Work’, ‘Education’ and, to a lesser extent, to the ‘Business’ activities. The further analysis of the ranked distance measures allowed identifying the following aspects:

- The activities ‘Bring Somebody’ and ‘Bring Something’ have the smallest distance value of all combinations (4.7) and the identical ranking order of distances to all other activities (the highest distance to Work and Education).

- The activity ‘Shopping’ shows - with one exception - exactly the same order of distances to other activities as the two ‘Bring Somebody/Something’ categories.
- There exists a very small distance value between the ‘Health’ and the ‘Service’ activities (7.0); there is no clear ranking order relating to the other activities; both activities show small distances only to ‘Other’ (9.8/14.3) and ‘Eat’ activities (16.2/13.8).
- ‘Visit’ and ‘Leisure’ also have a very small distance value (8.1); additionally, the five smallest distances values are to the same activities, nonetheless, in a different order.
- The category ‘Other’ does not show any clear assignment to the groups defined so far; almost the same distance is both to ‘Health’ (9.8) and ‘Leisure’ (10.0) and to ‘Visit’ (13.7) and ‘Services’ (14.3). Only the distance to the ‘Eat’ (8.6) activity is shorter.
- Category ‘Eat’ has the smallest distances to the ‘Bring Somebody/Something’ (11.7/10.8), ‘Shopping’ (8.2) and ‘Other’ (8.6) categories. At the same time, the ranking order of distances is very similar to those of the categories mentioned (except category ‘Other’).

For the final aggregation we have to take into the account that for some activities only a few observations are reported in the EOD. For instance, the ‘Business’ activity shows some of the highest distances to other activities and might be treated separately, but due to the minor number of reported mentions (1.5%) it has to be grouped with another activity. Notice that this final step of aggregating activities remains qualitative and somehow discursive, but the previous findings support the decision-making at this point.

- The ‘Work’ activity is merged with the ‘Business’ activity representing a common ‘Work’ activity. Although both categories differ by remarkable distances from all other activities, the ‘Business’ activity is the second nearest to ‘Work’ after the ‘Education’ activity. Another argument is that comparable travel destinations can rather be expected for the common ‘Work’ activity than for a ‘Business’ and ‘Education’ mixed category. The ‘Work’ activity total 21% of all reported activities/trips for the analysis.
- The ‘Education’ activity remains stand-alone as it shows - with only two exceptions - the second highest (after ‘Work’) distances to other activities and totals 12.2% of all activities/trips.
- The ‘Bring Somebody’ and ‘Bring Something’ activities (which show a high degree of similarity) are both assigned to the ‘Shopping’ activity, which they are comparable to. Additionally, the ‘Eat’ activity is assigned to this group as well because of the small distances. The group is denominated as ‘Shopping’ activity and covers 31.5% of all activities/trips.
- Another group consists of the ‘Visit’ and ‘Leisure’ activities. The analysis showed that no clear assignation to any other activity (or group) seems reasonable. Eventually, the question was if a grouping either with the ‘Health’ or with the ‘Service’ activities would have made sense. Nevertheless, according to the results as well as to our intention to not combine predominantly voluntary activities with those of a more compulsive character we have decided not to alter this group denominated as ‘Leisure’ with 18.1% of all trips/activities.

- Finally, a leftover group of activities denominated ‘Other’ is build. This group consists primarily of the activities ‘Other’ which represented in the analysis disperse behavior as no assignment to any other activity made sense. Additionally, the activities ‘Health’ and ‘Services’ are attached to the ‘Other’ activity. A combination with one of the other activities was not recommendable, and the two activities dispose together of only 8.1% of all activities/trips. At the same time, the ‘Other’ activity showed very small distances to both ‘Health’ and ‘Services’. Together, the so-called ‘Other’ group totals 17.2% of all activities/trips.

Summing up, the suggested grouping leads to five major activities of WORK (‘W’), EDUCATION (‘E’), SHOPPING (‘S’), LEISURE (‘L’) and OTHER (‘O’). However, we aggregated this groups even more, such that the ‘S’, ‘L’ and ‘O’ activities felt into a single representative category of secondary activities, denominated with ‘Y’. The groupings into five activities as well as the further aggregated version built the basis for the analysis of tour and activity pattern complexity described in 4.3.3 and illustrated in Figure 4.11.

Annex-2: Mode combination probabilities

Annex-Table 4: Mode combination probabilities by user groups

Mode	Mode Combination	no car / <= 150	1+ cars / <= 150	no car / > 150 & <= 300	1+ cars / > 150 & <= 300	no car / > 300 & <= 600	1 car / > 300 & <= 600	2+ cars / > 300 & <= 600
Public Transport	TpTpTp	0.455	0.455	0.460	0.457	0.562	0.530	0.529
	TpCaTp	0.345	0.182	0.347	0.261	0.223	0.230	0.294
	TpTpCa	0.091	0.000	0.094	0.130	0.081	0.020	0.000
	TpAaAa	0.009	0.273	0.015	0.043	0.035	0.070	0.118
	TpTpAa	0.036	0.000	0.045	0.043	0.065	0.080	0.000
	TpCaCa	0.055	0.000	0.020	0.000	0.019	0.040	0.000
	TpAaTp	0.009	0.091	0.020	0.065	0.015	0.030	0.059
Car Driver	AcAcAc	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Car Passenger	AaAaAa	0.500	1.000	0.647	0.906	0.708	0.894	1.000
	AaTpTp	0.500	0.000	0.294	0.000	0.208	0.064	0.000
	AaCaAa	0.000	0.000	0.059	0.094	0.083	0.032	0.000
	AaTpAa	0.000	0.000	0.000	0.000	0.000	0.011	0.000
Walking	CaCaCa	0.884	0.833	0.747	0.889	0.705	0.750	1.000
	CaTpTp	0.116	0.167	0.253	0.111	0.295	0.250	0.000
Bicycle	BiBiBi	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Mode	Mode Combination	no car / > 600 & <= 1.200	1 car / > 600 & <= 1.200	2+ cars / > 600 & <= 1.200	no car / > 1.200	1 car / > 1.200	2+ cars / > 1.200
Public Transport	TpTpTp	0.407	0.409	0.308	0.250	0.300	0.556
	TpCaTp	0.352	0.182	0.462	0.000	0.350	0.111
	TpTpCa	0.022	0.136	0.077	0.250	0.000	0.000
	TpAaAa	0.099	0.091	0.077	0.000	0.000	0.111
	TpTpAa	0.055	0.152	0.077	0.500	0.100	0.000
	TpCaCa	0.022	0.015	0.000	0.000	0.050	0.000
	TpAaTp	0.044	0.015	0.000	0.000	0.200	0.222
Car Driver	AcAcAc	1.000	1.000	1.000	1.000	1.000	1.000
Car Passenger	AaAaAa	0.700	0.815	0.960	0.500	0.926	0.912
	AaTpTp	0.200	0.074	0.000	0.250	0.000	0.000
	AaCaAa	0.100	0.074	0.000	0.000	0.037	0.059
	AaTpAa	0.000	0.037	0.040	0.250	0.037	0.029
Walking	CaCaCa	0.500	0.700	1.000	1.000	0.714	1.000
	CaTpTp	0.500	0.300	0.000	0.000	0.286	0.000
Bicycle	BiBiBi	1.000	1.000	1.000	1.000	1.000	1.000

Annotation: To be able to realize the analysis on more cases, the activity pattern types ‘HXYH’, ‘HYXH’ (with ‘X’ = work activity) and ‘HYYH’ were considered. / (N=1,891)

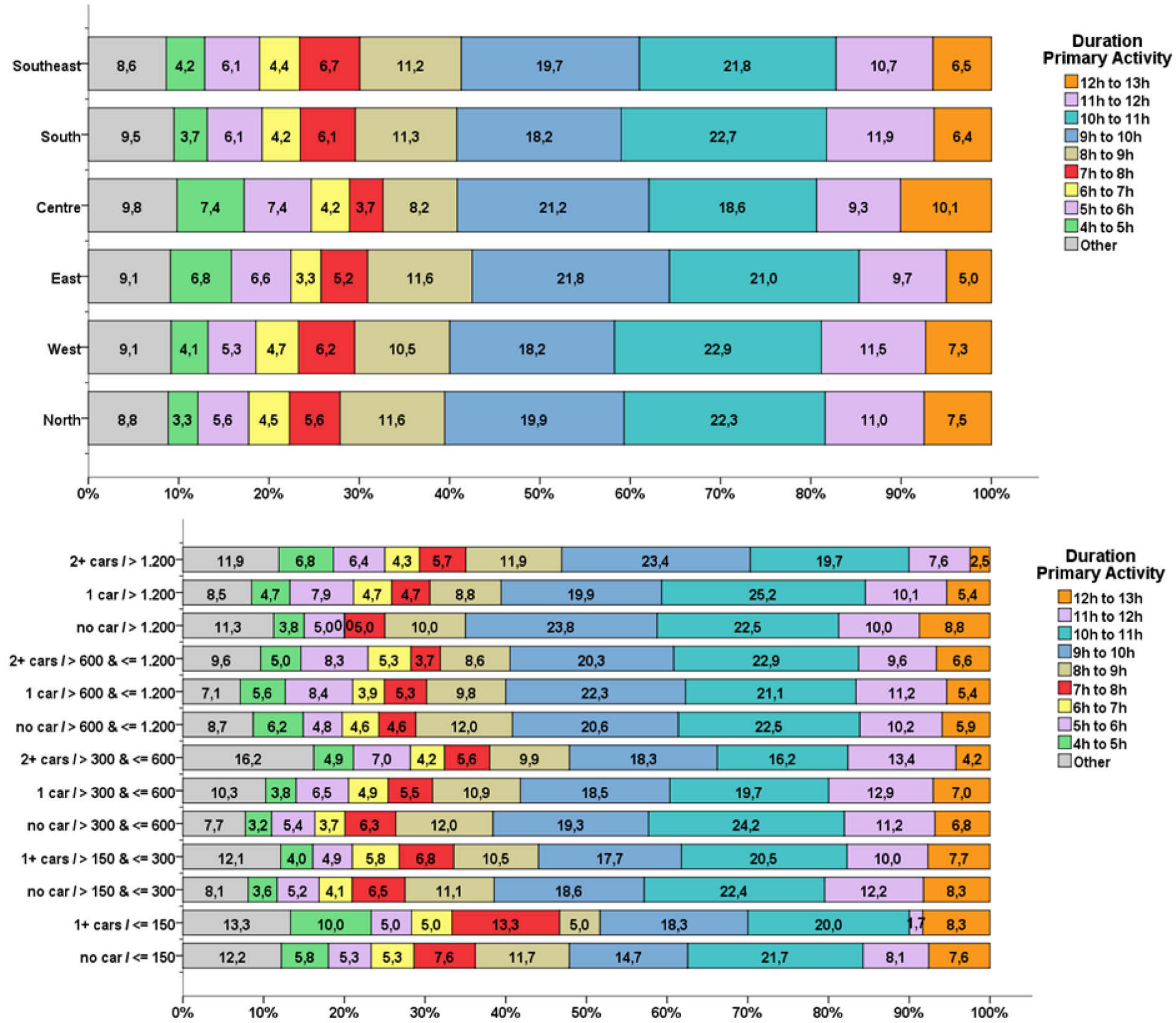
Annex-3: Final choice set

Annex-Table 5: Final choice set selection by mode combinations and tour relations

7		1		2		3		4		5		6		7		8		9		10		11		12		13		14		15	
		TpTpTp		TpCaTp		TpTpCa		TpTpAa		TpAaAa		TpCaCa		TpAaTp		AcAcAc		AaAaAa		AaTpTp		AaCaAa		AaTpAa		CaCaCa		CaTpTp		BiBiBi	
		HR	WR	HR	WR	HR	WR	HR	WR	HR	WR	HR	WR	HR	WR	HR	WR	HR	WR	HR	WR	HR	WR	HR	WR	HR	WR	HR	WR	HR	WR
		0.44	0.56	0.05	0.95	0.97	0.03	0.79	0.21	0.5	0.5	0.5	0.5	0.5	0.5	0.45	0.55	0.81	0.19	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Car Passenger																															
in %																															
1 to 10min	29.0%							2		2	2				2			2	2			2		2							
11 to 20min	30.6%							2		2	2				2			2	2			2		2							
21 to 30min	22.6%							2		2	2				2			2	2			2		2							
31 to 40min	8.1%							1		1	1				1			1	1			1		1							
41 to 50min	6.5%							1		1	1				1			1	1			1		1							
51 to 60min	1.6%							0		0	0				0			0	0			0		0							
> 60min	1.6%							0		0	0				0			0	0			0		0							
Car Driver																															
in %																															
1 to 10min	28.7%															2	2														
11 to 20min	25.5%															2	2														
21 to 30min	27.7%															2	2														
31 to 40min	4.3%															1	1														
41 to 50min	6.4%															1	1														
51 to 60min	2.1%															0	0														
> 60min	5.3%															1	1														
Walking																															
in %																															
1 to 10min	53.3%				4	4						4	4										4			4	4				
11 to 20min	31.7%				2	2						2	2										2			2	2				
21 to 30min	12.5%				1	1						1	1										1			1	1				
31 to 40min	.8%				0	0						0	0										0			0	0				
41 to 50min	.0%				0	0						0	0										0			0	0				
51 to 60min	.8%				0	0						0	0										0			0	0				
> 60min	.8%				0	0						0	0										0			0	0				
Public Transport																															
in %																															
1 to 10min	2.3%	0	0	0			0		0					0						1	1				1			1	1		
11 to 20min	21.8%	2	2	2			2		2					2						2	2				2			2	2		
21 to 30min	27.0%	2	2	2			2		2					2						2	2				2			2	2		
31 to 40min	15.5%	1	1	1			1		1					1						1	1				1			1	1		
41 to 50min	11.5%	1	1	1			1		1					1						1	1				1			1	1		
51 to 60min	14.9%	1	1	1			1		1					1						1	1				1			1	1		
> 60min	6.9%	1	1	0			1		1					1						1	1				1			1	1		

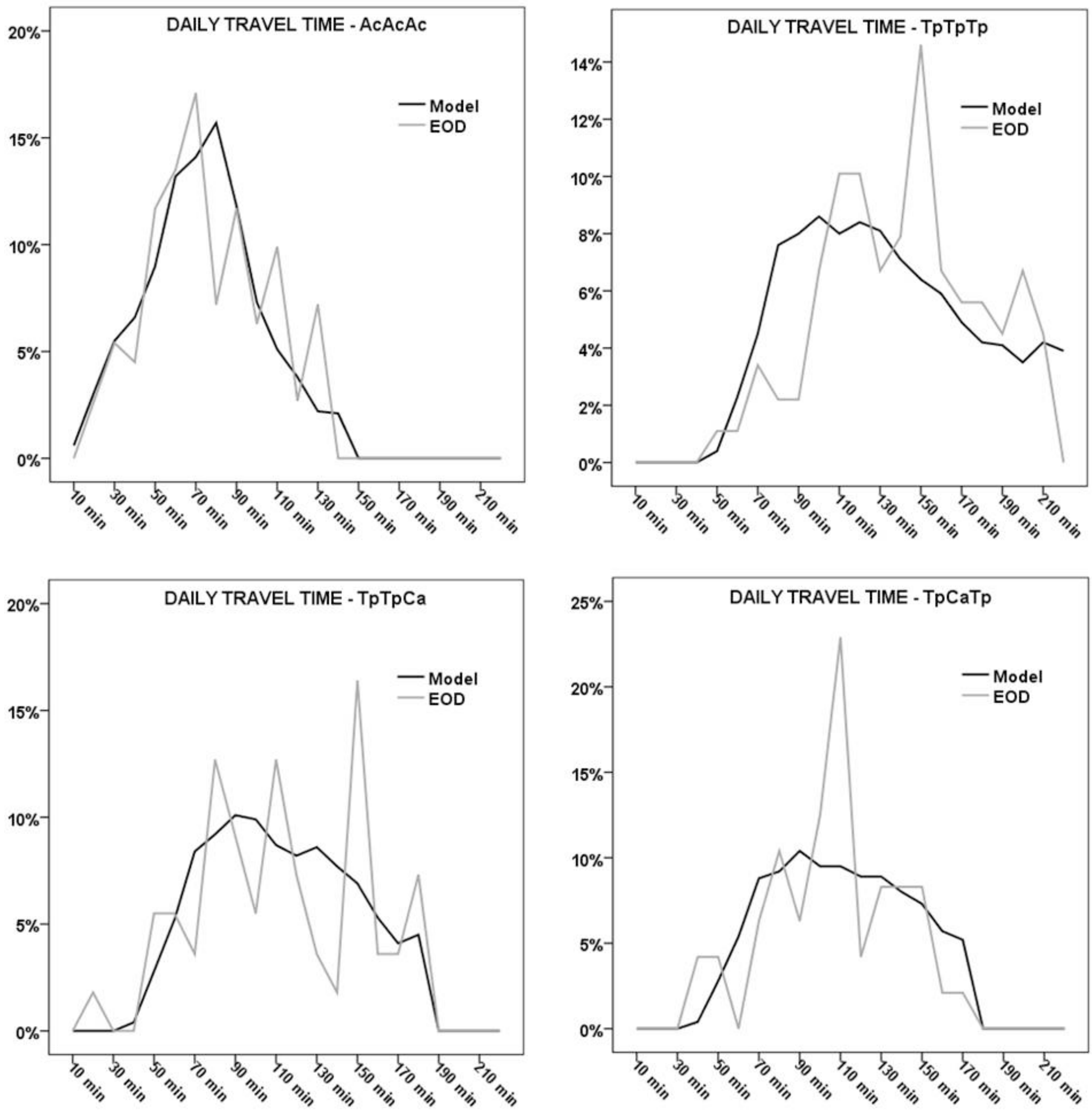
Annotation: In the upper row we differentiate between 15 mode combinations. Indicated below are the shares per mode combination, divided into home- (HR) and work-related (WR) tours. The values in columns reflect the number of alternatives selected by mode combination, tour relation and mode used towards the secondary location. Numbers in grey indicate alternatives that due to rounding would not have been selected (values below 0.5) but were included anyway.

Annex-4: Time regimes: duration of primary activity



Annex-Figure 1: Duration of primary activity by city sector and user groups

Annex-5: Daily travel times by mode combinations: model and EOD



Annex-Figure 2: Daily travel times by mode combinations – model and EOD

Annotations: EOD results are based on pattern types ‘HXYH’ and ‘HYXH’ with ‘X’ = work activity. In addition, time-space constraints active for the model (daily travel time thresholds and search spaces) were applied to the EOD data, too. / (AcAcAc, N=111; TpTpTp, N=89; TpTpCa, N=55; TpCaTp, N=48)

The peak-like behavior most probably can be explained by the tendency of respondents to round trip times. In general, we can conclude that tendencies of time consumption represented by the total daily travel time are reasonably reproduced with the model. The analysis is limited to the indicated mode combinations due to limited data available from the EOD for the other mode combinations considered in the model.

Digital Annex

Annex-Table 6: Overview of SPSS-Syntax used for EOD data mining

File name	Description
Syntax_EOD_Chain-Building.sps/txt	preparation and analysis of EOD trip file; generation of coherent tours and activity patterns
Syntax_PATTERNS+MODES.sps/txt	generation of the typology of activity pattern; analysis of daily travel times by mode combinations
Syntax_REFERENCES.sps/txt	analysis of tour relation (home- and work-related); trip time distributions towards secondary activity location
Syntax_Time.sps/txt	analysis of time regimes; correlation by city sectors and user groups
Syntax_HXYH_SpatialAnalysis.sps/txt	analysis of detour factors; regression models (based on file: HG_Viaje_Spatial.sav)
Syntax_PersonAttributes.sps/txt	definition of personal status variable and household type variable
Syntax_HouseholdAttributes.sps/txt	update of the households user group variable

Annotation: The digital annex contains both the original SPSS-Syntax as well as the same files saved in text-format. Respective steps of data mining can be examined via the text-files without running SPSS.

DVD-Folder: Data_Preparation

Annex-Table 7: Overview of SPSS-Syntax to run the model

File name	Description	Related files
SyntaxDemand_BASIS.sps/txt	represents major model calculation steps, including the activation of further syntax-routines	
SyntaxDemand_Durations.sps/txt	assignment of activities starting times and durations combinations	**TimeChoices.xls** (contains results of EOD analysis on time regimes)
WEIGHT_TT_TYP.sps/txt	integration of the mode combination probabilities	**WEIGHT_TT_TYP.xls** (contains results of EOD analysis and SPSS-Syntax)
WEIGHTING_SECONDARY.sps/txt	weighting of spatial path flows using LWPs	**WEIGHTING_SECONDARY.xls** (contains possibility to adjust LWPs and SPSS-Syntax)
SyntaxDemand_RANDOM.sps/txt	random process	
SELECT_SECONDARY.sps/txt	selection of the final choice set	**SELECT_SECONDARY_7.xls** (contains definition of the final choice set and SPSS-Syntax)
SyntaxDemand_CALIBRATION.sps/txt	calibration of the destination choice through the adjustment of the LWPs	

Annotation: The digital annex contains both the original SPSS-Syntax as well as the same files saved in text-format. Respective steps of model building can be examined via the text-files without running SPSS. The related files in Excel-format contain EOD analysis results and their transformation into SPSS-Syntax.

DVD-Folder: Model_Syntax

Annex-Table 8: Overview of the model’s input and output data files

File name	Description
HG_Viaje_02_2010.sav/txt	EOD data set; result of the application of Syntax: Syntax_EOD_Chain-Building.sps; coherent tours and activity patterns for Santiago
HG_Viaje_Spatial.sav	EOD data set (HG_Viaje_02_2010.sav); expanded by the straight-line distances for the home-work relation; basis for the analysis of detour factors (search spaces)
HG_Personas.sav/txt	EOD personal data file including personal status and household type variables
HG_Hogar.sav/txt	EOD household data file including updated household user group variable
USOSUELO_2002_Y_density.sav	land use density by TAZ provided by land-use model MUSSA
P_WORK_2007.sav	ESTRAUS matrix of the home-work relation by user group and transport modes
ALLCOSTS_FL.sav	ESTRAUS cost matrix (travel times by mode)
DURATIONS.sav	time regimes (starting times and durations) of activities 'X' and 'Y' by user group and origin
OUT_17.3.2011_70perc.sav	search spaces using different sets of detour factors (percentiles 60, 70 and 80); result of GIS-analysis
OUT_22.2.2011_60perc.sav	
OUT_31.3.2011_80perc.sav	
Orig_1to618_ScenarioTT.sav	model result; spatial path flows based on search spaces (percentile 70) - adjusted travel times
Orig_411to255_ScenarioTT_example.xls	model result, example of spatial path flows for one Home-Work relation; including extensive descpription of all modeling processes (based on Orig_1to618_ScenarioTT.sav)

Annotation: The digital annex contains the original SPSS-Data-Files. In addition, EOD trip, household and personal files are available in Excel-format, too. Due to its size, the final model file Orig_1to618_ScenarioTT.sav is not saved entirely in a distinct format. Nonetheless, the file Orig_411to255_ScenarioTT_example.xls outlines the complete model results along the example of one Home-Work relation. This file also includes short descriptions of all calculation steps.

DVD-Folder: Model_InputOutput

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides Statt, dass ich die vorliegende Arbeit selbstständig verfasst und ausschließlich unter Zuhilfenahme der angegebenen Quellen und Hilfsmittel verfasst habe. Ich versichere darüber hinaus, dass diese Arbeit in dieser oder einer anderen Form noch nicht anderweitig als Dissertation eingereicht oder veröffentlicht wurde.

Andreas Justen

Berlin, den 04. Mai 2011