Towards an activity-based modeling of transport demand in Santiago applying empirical schedules of transport behaviour

Megacities: Risk, Vulnerability and Sustainable Development

International Conference – September 6-8, 2009 / Leipzig

Contribution to the Young Researchers Workshop

Andreas Justen, DLR Institute of Transport Research (andreas.justen@dlr.de)

1. Central research questions

The central research questions are defined by the practice-oriented approach followed in the PhD thesis. The work presented here aims at the development and application of a transport model that allows for the impact assessment of transport policies and demographic scenarios. Respectively the very central research question can be summarized as follows:

“Does an activity-based analysis of transport demand in Santiago de Chile expand the possibilities of transport-policy analysis and can transport relevant processes be displayed both on a macro and a micro level[1]?”

In this context some further analytical questions are introduced:

- Are the data bases available in Santiago adequate to display the transport behaviour in a highly disaggregated manner of time and space?
- Can the activity-based approach be implemented by using a multi-scale approach to handle the complexity of the combinatorial number of options?
- How can activity-trajectories be modeled in a – time and space – consistent path and be adjusted to observed choices, e.g. regarding daily time budgets for activities and traveling?
- And finally, is it possible to transfer the methodology to other cities/regions?

2. Background and context

Transport models are today in many cases an indispensable component of urban transport planning. They support decision-making by the analysis of possible effects of infrastructure or transport management measures and hence, represent an integrative part of decision-making processes in urban planning units. Naturally there do exist also models that work on a nation-wide level but generally the models are applied on an urban or regional scale to display the interactions between transport supply (e.g. car infrastructure, public transport systems) and transport demand (behaviour of individuals and/or households). To this purpose models have to deal with complex urban environments on different spatial and temporal scales.

In Santiago de Chile exist a long tradition and experience with the application of a transport model in the urban planning and decision-making context. First applications go back to the development of the 4-Step-Model ESTRAUS since the 1980s [1,2]. Until today ESTRAUS is

[1] Further explanation of the terms “macro” and “micro” is given in the methodology-chapter.

regularly updated and applied to monitor the effects of fare policies or impacts of infrastructure projects. But meanwhile the requirements for the investigation of transportation demand evolve over the years; the modeling framework in Santiago remained nearly the same.

This is where the PhD thesis ties in. The work aims at developing a new analytical framework for the individual travel and activities demand specifying options in a multi-scale (time and space) hierarchical context such that makes the individuals search process more realistic and analytically feasible. The objective is to apply the framework to estimate Santiago's transport demand (daily chains of activities and trips), including mode and destination choices in the urban context. The paper starts with a short review of characteristics of current approaches and derives the need for their further development, delivers insight into the methodological concept under implementation and shows first results regarding the analysis of Santiago citizen's transport and activity behaviour. Finally an outlook is given regarding upcoming activities and working requirements.

3. Motivation

The well-established 4-step-models in transport planning are generally divided into four major steps of (1) the generation of trips, (2) their split in different transport modes, (3) the distribution of trips to destinations and (4) the assignment of all trips to road networks and public transport systems. Whereas the first and fourth step can be treated independently, the steps of mode and destination choices are handled either sequentially or in a combined manner. As the decisions for modes and destinations often depend on each other, recent approaches consider this by the estimation of joint mode-destination choice models. The same applies to Santiago's ESTRAUS model where the trip-generation step is followed by a conjoint and iterative step of partition, distribution and assignment until the systems equilibrium is achieved where (user) equilibrium refers to a state where no user of the system can achieve lower transportation costs by changing his or her decision.

Although the 4-step approach was successfully implemented in several commercial software packages (e.g. ESTRAUS, EMME/2, VISUM) [3] and can be applied by planners even without profound comprehension of the mathematical foundations, transport research entities enforced the development of more behaviourally sound approaches from the beginning of the 1990s [4,5,6]. All 4-step-models as well as ESTRAUS interpret travel demand as the aggregation and spatial distribution of trip flows meanwhile newer approaches understand travel demand as derived from interdependent activities and trips of individuals. In literature two main reasons are pointed out that motivated the appearance of activity-oriented approaches. First, as already mentioned, the shift from the trip flow perspective towards the analysis of individual's decision-making in transport, and second the paradigm change from supply-side oriented transport policies to those focused on the demand-side [7].[2]

Nevertheless, some specific characteristics of the activity-based approach have to be highlighted to clarify the differences. This is important regarding the general motivation background for research in the area as well as for the PhD thesis presented here. To start, (1) activity-patterns rather than individual trips are the unit of activity-based analysis, (2) resource constraints regarding time (e.g. daily individual time budgets for travel and activities, opening times of facilities) and financial resources may be considered, (3) household conditions rather than the individuals situation determine activity and travel behaviour and (4) coming back to the first point, activities, trips and the individuals socioeconomic characteristics remain connected throughout the analysis [8,5]. Another issue (5) which has often been addressed by activity-based approaches is the attempt to model transport behaviour on more detailed spatial units, e.g. on housing blocks or grid cells instead of using Traffic Analysis Zones (TAZ). To fulfill the requirements of an activity-based analysis of transport demand in Santiago the PhD thesis applies to some extent state of the art techniques but will focus on the provision of innovative methods to display transport related processes on a spatially differentiated bi-level and consider time-constraints within the analysis [9].

[2] As supply-side oriented measures the building of new car or public transport infrastructure serve as examples. Policies with focus on the demand-side are generally associated with those of pricing or taxation mechanisms and/or regulatory acts like parking prohibitions.

4. Methodology

The estimation of transport demand including all decisions, like when to start a trip, when to end an activity, when starting the next, which mode-destination combination to choose according to the mobility tools available and the individuals social background is not a trivial task, due to the huge amount of options and the information needed to analyze them. Respectively some basic assumptions of the framework presented here and its application to Santiago need to be introduced. A basic assumption concerning the methodology is that choice processes for activities and later on for destinations and modes are modeled as they were made hierarchically. This means for example that any activity-pattern (refers to the number and sequence of activities along one day) contains activities that are categorized by those of primary and secondary importance[3] [10]. Regarding the spatial hierarchy the individual's decisions are first made on an upper macro level and then further disaggregated to a micro level. In Santiago the macro level is characterized by 630 TAZ, the micro level consists of approximately 50.000 city blocks.

In general, several steps need to be addressed during the model setup:

1. *Population data set:* refers to the building of a data set with persons and households with multiple socio-economic attributes assigned to a spatial unit.
2. *Activity-patterns (transport behaviour):* means the extraction and analysis of individual activity-patterns using empirical schedules of transport behaviour and the assignment to members of the population data set.
3. *Decisions for modes and destinations (macro level):* corresponds to the estimation of joint probabilities for mode-destination choices in accordance to disaggregated behavioural-homogeneous groups identified in the population data set.
4. *Timing:* refers to the assignment of probabilities of starting times and durations of activities to adjust the joint probabilities against constraints (time, costs).
5. *Decisions for modes and destinations (micro level):* implies the estimation of activity spaces based on empirical observations and disaggregated land-use data to adjust probabilities for secondary activity choices.

Step 1 and the building of the *population data set* is pending whereas methodologies will comprise standard techniques like the calculation of general statistics (frequencies, cross tables) of publicly available data sources, as well as regression models, e.g. to estimate car or driving license ownership. The method of Iterative Proportional Fitting (IPF) will be applied

[3] In activity-based analysis the term "primary" refers to activities of compulsive character like Work or Education,; where "secondary" refers to activities with a discretionary notion.

to combine aggregated demographic data of persons and households with disaggregated information in form of multi-dimensional cross-tables about socio-economics of persons and households.[4] The IPF procedure serves to adjust iteratively the internal structure of a cross-table to given row sums [11,12].

At the time of writing step 2, 3 and 4 are the most advanced ones with a completed methodology to generate and analyze activity-patterns. Step 2 initially required an *activity classification* (method do group similar activity types), an *activity prioritization process* (corresponds to the determination of primary and secondary activities) and a rule-based assignment of single patterns into *activity-pattern-groups*. The following chapter about first results clarifies the methods applied for step 2.

Sources to the definition of *joint probabilities of mode and destination choices* in step 3 are two-fold. Probabilities for primary activities are taken in accordance to the information given by the ESTRAUS output in form of Origen-Destination (OD) matrices. This is done considering the concept of hierarchical decision-making as location and mode choice of the primary activity determines adjacent processes. ESTRAUS divides the transport demand into 13 user categories[5] (households by income and number of cars), 3 activities (work, education, other) and 6 modes (car driver, car passenger, public transport, walk, taxi, shared taxi). Hence, in a first attempt we use the differentiation of 234 OD matrices and apply the probabilities to primary activities in each activity-pattern. This means, if for example an individual of the population data set that belongs by its socio-economic attributes to the ESTRAUS user category 4, living in TAZ X and applies an activity-pattern which includes the activity of “Work”, this pattern gets assigned the probability array of the calculated ESTRAUS mode-destination combination (see Matrix 5). To get a better understanding of how the information of the ESTRAUS model is used, see the following matrix examples.

	Sum	1	2	3	4	5	6
	19273.622	12.737	45.931	36.063	47.543	17.514	34.2
1	12.772	0	0	0	0.614	0.410	1.2
2	49.677	0	0	0	0.658	0.524	1.4
3	24.907	0	0	0	0	0	
4	30.375	0	0	0	0	0.446	1.0
5	16.607	0	0.306	0	0	0	0.4
6	29.667	0	0.383	0	0.297	0	

Matrix 1: ESTRAUS OD matrix, user category 04, purpose “Work”, mode “car driver” (ach)

[4] The demographic data about number of persons and households is given by the Chilean National Census of 2001 and is updated using current statistics from the National Statistics Institute (INE); the information about household structure (single and family-households) and socio-economics (especially household income) is taken from the National Characterization Socio-economic Survey (CASEN).

[5] The user categories of ESTRAUS are as follows: **01** / 0 cars, 0-148.226 CHP, **02** / 1+ cars, 0-148.226 CHP, **03** / 0 cars, 148.227-296.452 CHP, **04** / 1+ cars, 148.227-296.452 CHP, **05** / 0 cars, 296.453-592.904 CHP, **06** / 1 car, 296.453-592.904 CHP, **07** / 2+ cars, 296.453-592.904 CHP, **08** / 0 cars, 592.905-1.185.808, **09** / 1 car, 592.905-1.185.808, **10** / 2+ cars, 592.905-1.185.808, **11** / 0 cars, > 1.185.808 CHP, **12** / 1 car, > 1.185.808 CHP, **13** / 2+ cars, > 1.185.808 CHP

	Sum	1	2	3	4	5	6
	40023.248	29.926	90.488	83.826	94.188	35.310	63.9
1	25.018	0	0	0	1.295	1.165	2.5
2	95.501	0	0	0	1.428	0.987	2.6
3	46.869	0	0	0	0	0	
4	58.231	0	0	0	0	0.862	2.8
5	33.479	0	0.579	0	0	0	1.2
6	56.646	0	0.747	0	0.793	0	

Matrix 2: ESTRAUS OD matrix, user category 04, purpose "Work", all modes

	Sum	1	2	3	4	5	6
	1442135.975	714.229	4961.738	1068.210	4146.644	2623.258	2044.5
1	809.166	0	0	0	47.398	35.145	49.1
2	2816.299	0	0	0	46.080	53.145	53.8
3	1124.760	0	0	0	0	0	
4	2111.564	0	0	0	0	51.788	37.1
5	2366.061	0	52.926	0	0	0	37.2
6	2853.359	0	51.266	0	37.434	0	

Matrix 3: Share of car driver mode (ach) by OD pairs, user category 04, purpose "Work"

The share of modes by OD pairs is result of the division of Matrix 1 by Matrix 2 (multiplied by 100), see e.g. the OD pair 1 to 4 (blue colored cells)

	Sum	1	2	3	4	5	6
	61300.000	36.711	152.370	83.893	168.882	57.542	108.5
1	100.000	0	0	0	5.178	4.658	10.2
2	100.000	0	0	0	1.495	1.033	2.8
3	100.000	0	0	0	0	0	
4	100.000	0	0	0	0	1.480	4.8
5	100.000	0	1.730	0	0	0	3.8
6	100.000	0	1.318	0	1.400	0	

Matrix 4: Share of trips for each OD by row, user category 04, purpose "Work"

The share of trips is result of the division of each matrix entry by the row sum; see Matrix 2 and the calculation of 1.295 (blue colored cell) / 25.018 (row sum) = 5.176.

	Sum	1	2	3	4	5	6
	30019.629	18.508	80.665	40.880	86.894	27.941	57.4
1	51.052	0	0	0	2.454	1.637	5.0
2	52.017	0	0	0	0.689	0.549	1.5
3	53.141	0	0	0	0	0	
4	52.162	0	0	0	0	0.767	1.8
5	49.605	0	0.915	0	0	0	1.4
6	52.372	0	0.676	0	0.524	0	

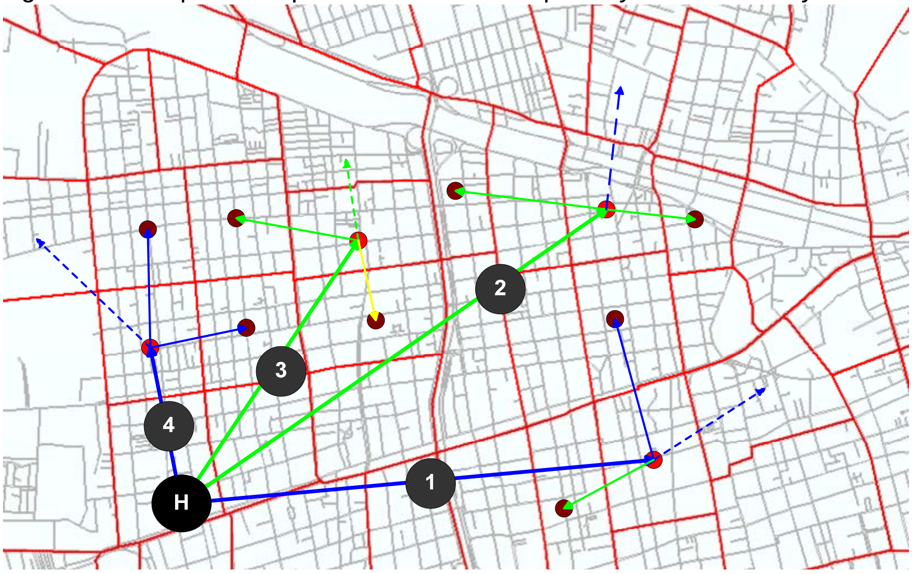
Matrix 5: Probability that an OD pair is realized as car driver, user category 04, purpose "Work"

The row sum value of 51.052 indicates that this share of trips coming from TAZ 1 is realized as car driver whereas the share of 2.454 of all trips ends up in TAZ 4 (Matrix 4 * Matrix 3) / 100. Matrix 5 can be calculated also by the division of each value of Matrix 1 by the row sum of Matrix 2 (e.g. (0.614 / 25.018) * 100). Nevertheless the calculated destination probability of Matrix 4 may be used later within the analysis.

Anyhow, choices for secondary activities along an activity-pattern are derived in the first version of the model by using empirical and static OD probabilities Santiago's Origin-

Destination-Survey (ODS) provides [13].[6] The Survey includes geographical coordinates for all trips and allows for the building of OD matrices differentiated by mode. Likewise it is possible to estimate mode transition probabilities directly from the ODS according to the type of activity-pattern executed. As a result joint probabilities for mode-destination combinations can be calculated. Even though a huge range of spatial activity-trajectories (refers to an activity-pattern complemented by destinations and modes) can be estimated, not all of them seem to be viable due to network impedance values (costs, times) and in accordance to the travelers' time restrictions. This is when the estimated joint probabilities have to be checked against their compatibility of criteria like daily travel and activity time budgets. The following figure visualizes the so far introduced methodology for the estimation of joint probabilities.

Figure 1: Example for dependencies between primary and secondary choices



Annotation: The red zones refer to Santiago's TAZ structure; the underlying disaggregated zones represent the city-block level. In this simplified example based on the home-location "H" (which is geocoded on the city-block level) exist four probabilities for primary destinations (red spots), with two OD pairs by public transport (green lines 2 and 3) and two OD pairs by private car (blue lines 1 and 4). These probabilities are taken from the ESTRAUS model output (see Matrix 5). Any red spot refers to the geographical center of a TAZ and subsequent OD probabilities for secondary activities (arrows to brown spots) are calculated directly from the ODS and in accordance to the zone and activity-pattern type executed. The interrupted arrows stand for initially feasible joint probabilities that are not considered any more due to criteria of time constraints.

Step 4 is about the integration of *timing aspects* in the calculation of activity-trajectories. Due to the probabilistic approach the question is where the distributions of starting times and durations of activities should be considered. The proposal is once again two-fold: first, the probability that an activity takes place in any TAZ (see Matrix 4) may be further disaggregated by combined probabilities of starting times and durations (see Figure 5). In this case the value of 5.178 (Matrix 4, probability of the OD relation 1 to 4) is multiplied by the

[6] Santiago's Travel Survey (Encuesta Origen Destino 2001) provides information about approximately 150,000 trips as well as sociodemographic and economic characteristics of nearly 60,000 people living in 15,000 households.

probability that e.g. the activity starts between 7h and 8h and takes 5h.[7] Now, for any (timely disaggregated) activity-trajectory the costs in form of time and/or financial expenditures can be calculated using impedance matrices the network model delivers.[8] In the case of complex activity-patterns with several secondary activities the impedances for each OD relation have to be added. Subsequently, expected times/costs for any activity-trajectory are mirrored against a "typical" daily travel time budget for the activity-pattern-group (see Table 4). If the budget is heavily exceeded the probability for respective activity-trajectories ought to decrease. The development of an approach to adjust the probabilities of activity-trajectories is not trivial as the standard deviation of time budgets is high and the most relevant variables that determine the budget need to be identified. It is already evident that the pattern-length and thus the number of trips per day have influence on the daily travel time budget (see Table 4). Further investigation is necessary to check for the influence of other variables, such like socio-economics, car availability or residential area.

Step 5 *(decisions for mode-destination combinations on the micro level)* is of special interest within the PhD thesis as it aims at a further refinement of the so far presented approach. The spatiotemporal choices for any complete activity-trajectory are taken at the macro (TAZ) level; nevertheless "zooming" into the micro level will be possible for selected TAZ. For this purpose trips arriving at a certain TAZ (e.g. the red and brown spots in Figure 1) are further distributed to the micro level according to the used mode, the land-use characteristics, the transport supply quality in the area and the socioeconomics of travelers. Attempts have already been underway to analyze the influence of urban form on transport behaviour in Santiago, but missing and less disaggregated data did not allow achieving satisfactory results [14]. Regarding the framework presented here, the consideration of processes on the micro level aims at the replacement of the initial and static OD distributions given by the ODS (see above). Therefore, the influence of respective variables on distances or activity spaces traveled and modes chosen is determined applying (prospectively) a multinomial logit model. The goal is to provide a higher solution of the respective processes (trips and activities) for a restricted amount of TAZ. Although the estimation of the respective choice model is pending, the already available, acquired and analyzed data for Santiago is quite promising.[9]

[7] Referring to Figure 5 the probability is calculated by the division of the number of cases in the respective square-combination of starting time and duration by the overall number of observations.

[8] ESTRAUS contains the networks for road infrastructure as well as bus and metro lanes. Based on that generalized costs for any OD pair are calculated.

[9] For each trip reported in the ODS the geographical coordinates on the micro level are available. As shown in Figure 1trip distances and activity-spaces can be measured.

5. First results

Regarding first results the analysis of the OD Survey aiming at the provision of activity-patterns as input to the model is completed. The ODS differentiates 12 activities that needed to be aggregated, primarily due to the limited number of observations for some activities. The empirical data of the ODS is not only used to build activity-patterns and OD distributions but to deliver information about the distributions of starting times and durations of activities that are leaked to the patterns. The *activity classification* is done hierarchically considering first the statistics of both the activity and trip durations, second the modal split values and third the frequencies in which activities occurred (the following explanations refer to Table 1).

Table 1 : Transport characteristics by 11 disaggregated ODS activities

Activity Type		Activity Duration	Trip Duration		car driver	car passenger	public transport	bicycle	walking	shared taxi	taxi	other
Work	Mean	476	40	MODAL SPLIT	19,4	6,9	49,7	3,8	14,4	2,3	0,7	2,7
	Median	530	33									
	Standard Deviation	203	31									
	Valid n	14.413	14.413									
	Share of purpose	19,82	19,82									
Education	Mean	341	25	MODAL SPLIT	2,0	13,6	35,8	0,6	33,8	1,2	0,4	12,6
	Median	325	20									
	Standard Deviation	126	23									
	Valid n	9.040	9.040									
	Share of purpose	12,43	12,43									
Shopping	Mean	43	15	MODAL SPLIT	9,7	8,8	14,7	1,6	62,0	2,2	0,6	0,3
	Median	20	10									
	Standard Deviation	60	28									
	Valid n	16.182	16.182									
	Share of purpose	22,26	22,26									
Visit	Mean	198	26	MODAL SPLIT	11,2	17,2	26,6	3,0	37,6	1,7	1,4	1,2
	Median	150	18									
	Standard Deviation	166	29									
	Valid n	7.720	7.720									
	Share of purpose	10,62	10,62									
Other	Mean	123	23	MODAL SPLIT	5,0	22,3	23,1	0,8	42,9	3,3	1,5	1,1
	Median	75	15									
	Standard Deviation	146	32									
	Valid n	6.732	6.732									
	Share of purpose	9,26	9,26									
Services	Mean	73	30	MODAL SPLIT	12,1	6,2	49,9	2,0	21,8	5,8	1,3	0,9
	Median	45	25									
	Standard Deviation	82	23									
	Valid n	4.097	4.097									
	Share of purpose	5,64	5,64									
Leisure	Mean	156	23	MODAL SPLIT	8,1	15,6	17,3	4,4	51,1	1,4	1,1	1,1
	Median	120	15									
	Standard Deviation	129	37									
	Valid n	5.645	5.645									
	Share of purpose	7,76	7,76									
Health	Mean	104	31	MODAL SPLIT	6,0	11,1	51,3	0,3	19,8	6,2	4,2	1,2
	Median	80	25									
	Standard Deviation	95	22									
	Valid n	1.881	1.881									
	Share of purpose	2,59	2,59									
Bring somebody	Mean	22	20	MODAL SPLIT	30,4	11,1	11,3	1,4	43,0	1,2	1,0	0,5
	Median	5	15									
	Standard Deviation	65	28									
	Valid n	5.057	5.057									
	Share of purpose	6,96	6,96									
Eat	Mean	82	16	MODAL SPLIT	15,1	14,4	10,6	1,8	55,1	1,0	1,3	0,6
	Median	58	10									
	Standard Deviation	70	39									
	Valid n	1.260	1.260									
	Share of purpose	1,73	1,73									
Bring something	Mean	41	19	MODAL SPLIT	20,7	12,4	17,0	4,6	39,5	2,8	1,9	1,0
	Median	15	15									
	Standard Deviation	73	20									
	Valid n	676	676									
	Share of purpose	0,93	0,93									

Source: ODS, 132.986 trips in 41.970 patterns

The activities Work and Education are not further aggregated as they show by far the highest mean values for activity duration, are also interpreted individually by the ESTRAUS model and represent a major amount of cases. The high standard deviation for the Work activity duration indicates a remarkable share of part-time jobs included that explain the differences between the lower mean and higher median value. Shopping, due to the high share of activities could be analyzed as stand-alone activity; anyhow the activity Eat shows very similar, low values for the trip duration. In combination with the highest share in walking of all activities it seems that both activities are conducted primarily within short walking distances. Visit and Leisure show (after work and education) the longest average activity durations and also similarities in trip durations whereas the standard deviation of the activity duration indicates a widespread timely distribution over the day. Health and Services show conformity especially as the average activity durations are in-between the long time activities of Work/Education and the short time activities of Shopping or the pick up and delivery activities (Bring Somebody, Bring Something). Trip durations are the highest after commuting times and there is a strong conformity along the by far dominant mode share of public transport and walking. The leftovers-category Other comprises a considerable share of activities, a very high standard deviation for the activity duration and a modal split that does not fit well either to the Health/Services nor Visit/Leisure categories. The categories of Bring Somebody and Bring Something are aggregated despite some variations in the activity duration statistics due to evident similarity in trip characteristics and the highest share of car driver mode.

After the first step of aggregation neither the Bring Somebody/Bring Something nor the Health/Services or Other activity category represent a satisfying number of trips and activities. Nevertheless the Visit/Leisure category is not further combined with one of the other categories to avoid the mixture with activities that include certain compulsive elements (e.g. Services, Other). Hence, due to the intention that any activity category should cover an appropriate amount of trips/activities comparable the Work/Education/Shopping categories and the impossibility to treat one of the categories as stand-alone activities, we assign the remaining categories to the leftover-category Other. Thus, the following Table 2 shows the characteristics of the 5 aggregated activity-types that build the background for further analysis.

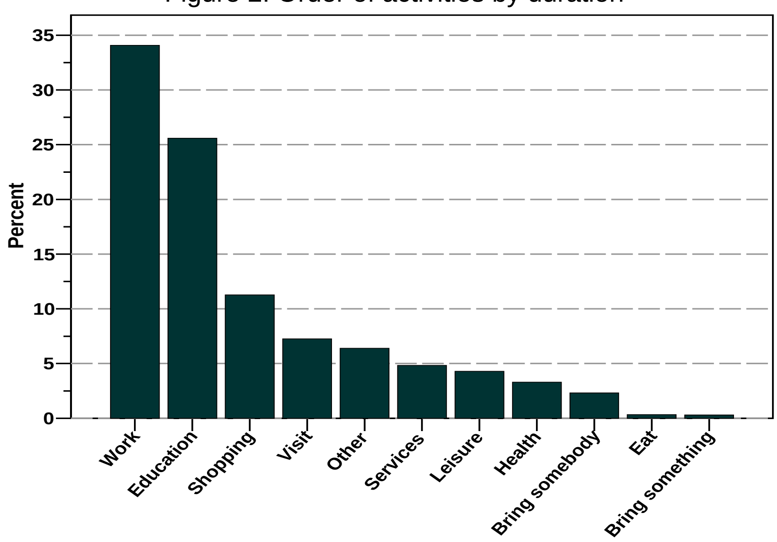
Table 2: Characteristics of 5 aggregated ODS activity types

Activity Type		Activity Duration	Trip Duration		car driver	car passenger	public transport	bicycle	walking	shared taxi	taxi	other
Work	Mean	476	40	MODAL SPLIT	19,4	6,9	49,7	3,8	14,4	2,3	0,7	2,7
	Median	530	33									
	Standard Deviation	203	31									
	Valid n	14.413	14.413									
	Share of purpose	20	20									
Education	Mean	341	25	MODAL SPLIT	2,0	13,6	35,8	0,6	33,8	1,2	0,4	12,6
	Median	325	20									
	Standard Deviation	126	23									
	Valid n	9.040	9.040									
	Share of purpose	12	12									
Shopping	Mean	46	15	MODAL SPLIT	10,1	9,2	14,4	1,7	61,5	2,1	0,7	0,3
	Median	25	10									
	Standard Deviation	62	29									
	Valid n	17.442	17.442									
	Share of purpose	24	24									
Leisure	Mean	180	25	MODAL SPLIT	9,9	16,6	22,7	3,6	43,3	1,6	1,3	1,1
	Median	135	15									
	Standard Deviation	153	33									
	Valid n	13.365	13.365									
	Share of purpose	18	18									
Other	Mean	79	24	MODAL SPLIT	14,2	14,2	28,5	1,3	35,8	3,5	1,6	0,9
	Median	40	17									
	Standard Deviation	116	28									
	Valid n	18.443	18.443									
	Share of purpose	25	25									

Source: ODS, 132.986 trips in 41.970 patterns

Based on the activity classification in the next step activity-patterns were extracted from the ODS and primary activities defined. Sometimes the concept of *activity prioritization* – that truly helps to structure processes and reduce complexity – is set ad-hoc making assumptions without checking for empirical evidence. To identify activities of primary importance and proof the assumption that Work and Education activities define daily travel behaviour, for every pattern the activity with the longest duration was labeled [10]. Hence, we check if the individual's major daily time investment reflects the order of the "assumed" activity importance. To get a broader picture of the ordering, Figure 2 considers all 11 activities for the analysis.

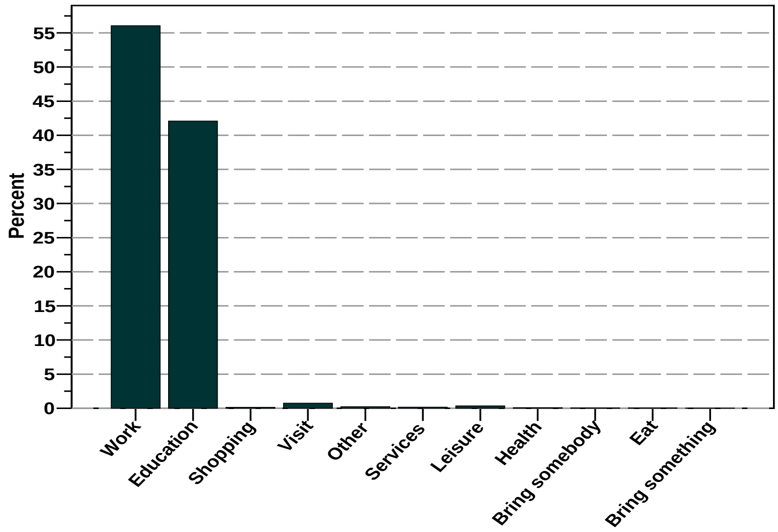
Figure 2: Order of activities by duration



Source: ODS, workday (Monday-Friday), 31.597 patterns

The results indicate the relation between time resource consumption and the compulsive activities of Work and Education (59.6% of all patterns). If we push this analysis forward and select all patterns that contain a work or education activity and apply the criteria of the longest duration, we can conclude that in almost every case when a Work or Education activity appears within a pattern it is the most important activity of the day (98.2%, Figure 3).[10]

Figure 3: Type of primary activity in pattern with work or education activity



Source: ODS, workday (Monday-Friday), 19.209 patterns

The setting of Work and Education as primary activities is then used to build categories of pattern-types. Following the idea that the hierarchies of primary and secondary activities determine subsequent decisions we consider this when building the categories. Basically this means for example to distinguish between pattern that contain a primary activity (Work/Education) and those that consist exclusively of secondary activities (Leisure, Shopping, Other). We define rules according to the recent findings of primary and secondary activities that enable the assignment of each pattern into one of 3 main-categories and 20 sub-types (see Table 3)[11].

[10] The picture gets much more heterogeneous when we analyze (the opposite) patterns without any work or education activity regarding the longest duration the remaining activities. In 28.5% of all patterns Shopping was the longest activity, in 17.4% Visit, in 15.9% Other, in 12.1% Services and in 10.4% Leisure (n = 12.388 patterns).

[11] Before the categorization manifold rules were applied to the ODS data set to clean and standardize the activity patterns. This refers to conditions such like that any pattern has to start and end at the home-location, that no primary activity may occur consecutively or that patterns with missing or wrong data about times were removed from the data set.

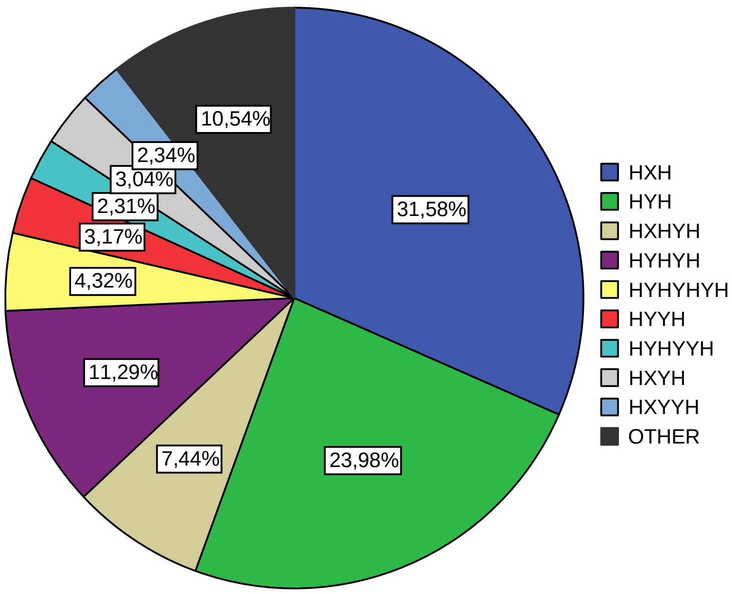
Table 3: Activity-Patterns Categories

Pattern Category	Abbreviation	Examples	Description
HOME-BASED-SINGLE-TRIPS	HBSINT	HXH HYH HXHYH HYHYHXH HXHXHYHYH	Refers to an independent number of consecutive trips whereas every second activity takes place at the home-location. In other words: two consecutive out-of-home activities never occur. The activities between two home-stays may be of primary or secondary importance.
NON-HOME-BASED SECONDARY TOURS	NHBSECT	HYYH HYHYYH HYYYH HYYHYYH HYHYYHYYH	In this category only activities of secondary importance appear. The category is labeled non-home-based as two or more activities of secondary importance can occur consecutively.
NON-HOME-BASED TOURS	NHBT	HXYH HYXH HXYXYH HXHYYH HYHXHYYH	Unlike the NHBSECT a primary activity occurs at some point of the pattern. Another characteristic is that two consecutive activities without a home-stay in between necessarily occur; otherwise the pattern would belong to the category of HBSINT.

Annotation: The letter “X” represents a primary activity (Work/Education), the letter “Y” represents any secondary activity (“H” stands for Home). The column with examples of activity-patterns is no exhaustive enumeration of different types but gives an idea of what kind of combination is possible within each category. This means that within the displayed sub-type HXHYYH appear also sequences of HYYXYH or HYHYXH. The 20 sub-types emerge because of different pattern-lengths in each pattern category, thus we define 8 sub-types for the HBSINT-category and 6 sub-types for each of the NHBSECT and NHBT categories.

The frequency distribution of the pattern categories indicates a strong concentration of HBSINT pattern within the ODS data set (81.8% of all pattern). The following Figure 4 summarizes the pattern sub-types and their frequencies.

Figure 4: Frequencies of activity-pattern sub-types



Source: ODS, 41.970 patterns

Annotation: The legend indicates examples for the respective pattern sub-types as explained in Table 3. The legend item “OTHER” comprises all sub-types that occur with a share of less than 2%.

Regarding the implications for model building we can summarize that the great majority of patterns contain simple home-based single trips, followed by 9.2% of non-home-based secondary tours and 9.0% of non-home-based tours. This implies that by focusing on the home-based patterns the great majority of the overall observed transport behaviour in Santiago can already be reproduced.

In the methodology-chapter we addressed the necessity to consider *timing aspects* within the framework. The ESTRAUS model delivers OD matrices for the morning-peak hour but misses to describe the overall timely distribution of the transport demand along the day. Hence, the first methodological aspects regarding the integration of time requires for information about constraints and distributions of time use. The following Table 4 gives an impression about the parameter "daily travel time budget" that is included in the adjustment process.

Table 4: Statistics of the daily travel time budget by pattern sub-types

		daily travel time budget		
		Valid n	Mean	Standard Deviation
pattern sub-type	HXH	12184	83	60
	HYH	5634	56	57
	HXHYH	2783	96	64
	HYHYH	2864	71	58
	HYYH	783	96	60
	HYYYH	322	120	63
	HYHYYYH	231	141	73
	HXHYHXYH	236	137	62
	HXHYYHYYH	117	142	96

Source: ODS, workday (Monday-Friday), 31.597 patterns (table with a selection of 25.154 patterns)

The daily travel time budget reflects the summed travel times over one day. The statistics for a selected number of pattern sub-types indicate that the time spent in traveling depends – as expected – on the number of trips realized. Nonetheless there is no linear correlation between the number of trips and the time spent in traveling. We can carefully conclude that the longer the pattern gets, the shorter the trips get (in time). If we check time expenses for each trip along a pattern for its mean values we discover that the average time spent per trip decreases considerably between the second and third trip and then declines continuously (see Table 5).

Table 5: Average travel time by 8 trips

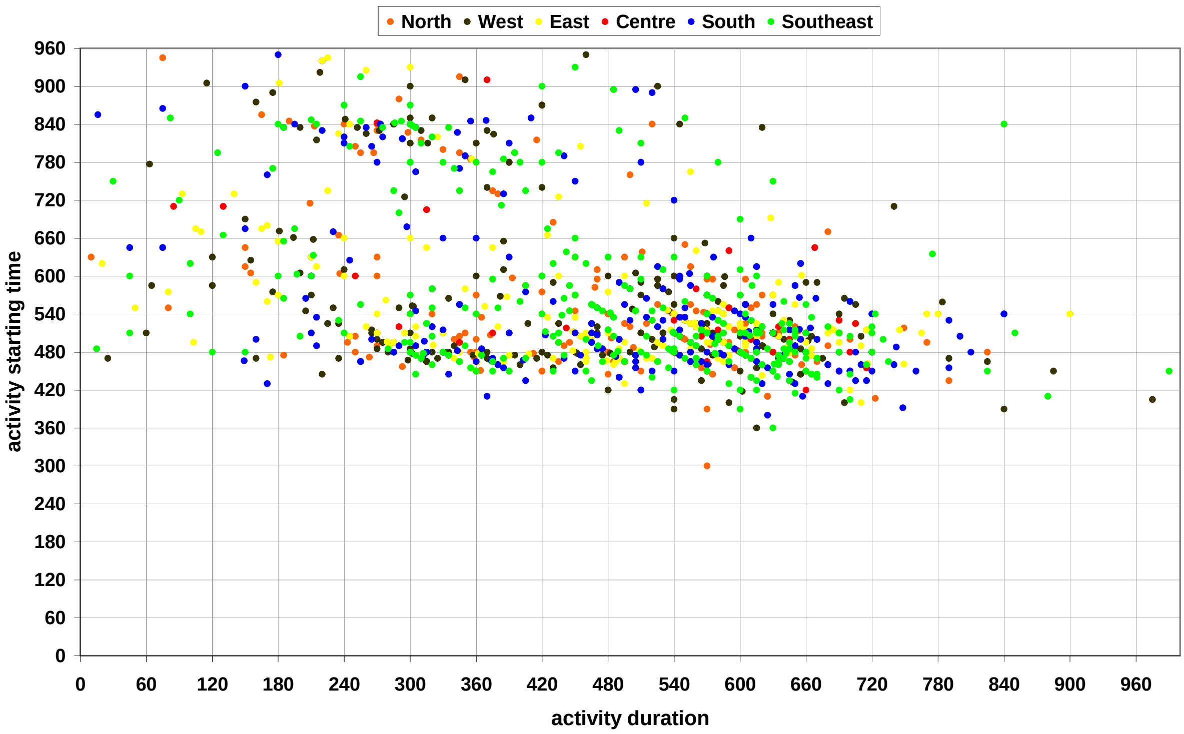
duration trip 1		duration trip 2		duration trip 3		duration trip 4	
Count	Mean	Count	Mean	Count	Mean	Count	Mean
31597	31	31597	32	13732	21	11789	21

duration trip 5		duration trip 6		duration trip 7		duration trip 8	
Count	Mean	Count	Mean	Count	Mean	Count	Mean
4911	19	3675	17	1277	15	708	14

Source: ODS, workday (Monday-Friday), 31.597 patterns

The second important information regarding a timely differentiation of activity-patterns is about the starting and duration times of activities. To exemplify the information that can be provided by the analysis of the ODS we select the pattern sub-type HXYH and display the starting times and durations of the primary activity (see Figure 5).

Figure 5: Combination of activity starting times and durations by Santiago's city sectors



Annotation: Every position of a point reflects the combination of starting time (Y-axis) and duration of the primary activity (X-axis) (additionally differentiated by Santiago's city sectors). Both axis display the time of day in minutes, e.g. the position (X=540/Y=480) means the activity started at 8h in the morning and took 9 hours. N = 1.274.

The figure allows for the detection of three rough clusters. The first and most concise one with primary activities starting between minute 420 (7h) and 600 (10h) taking between 420 and 720 minutes (e.g. full-time jobs), the second one with similar starting times but durations of 240 to 420 minutes (e.g. part-time jobs, morning) and third, activities starting later at minute 780 (13h) to 870 (14h30) taking 180 to 360 minutes (part-time jobs, afternoon). The implications for the model are two-fold: first, it needs to reproduce the general time-structures referring to the basic behavioural elements of for example full-time and part-time working activities, second it needs to be flexible to represent also the variances in behaviour that do not fit into one of the observed clusters.

Summing up, until this point of the research the first results demonstrate that both in the case of providing activity-patterns as well as regarding the information about timely distributions of activities, the required information for activity-based analysis is given. The methodology for the calculation of joint-probabilities on the macro-level also seems quite clear and advanced substantially; however, the extent to which processes may be modeled at a disaggregated level remains currently conditioned by data quality and availability.

6. Perspectives and outlook

The presented methodologies and first results necessarily have to be understood as insight into a work in progress that still experiences changes and adaptations. Nevertheless the basic structure to model transport behavioural processes in Santiago was introduced. The main focus in the upcoming months is about the implementation of timing constraints in a first version of the model as well as the estimation of activity spaces for secondary choices on the micro level as this reflects a major innovative aspect of this work. Another issue – as soon as a first version is applicable – is about the selection of transport-policies to test the model against its abilities to respond to measures. Anyhow, as the PhD thesis aims at an introduction of a whole new framework for the analysis of transport in Santiago, not every step is developed with utterly new methodologies. However, innovation in the area can be seen in the introduction of a spatial bi-level for the analysis of transport demand and the disaggregation of the demand in time. Additional value of this work refers to its comprehensiveness itself as it tackles every step of transport demand generation and thus tries to implement a new modeling philosophy in Santiago. Its importance beyond the case study of Santiago will be measured against its applicability by third parties and the possibility to transfer methodological elements to other cities or regions.

7. References

[1] Ministerio de Planificación (2003): ESTRAUS. Manual del Usario. Version 4.3. Santiago, Chile.

[2] Florian, M., Wu, J.H., He, S. (1999): A multi-class multi-mode variable demand network equilibrium model with hierarchical logit structures. Montreal, Canada.

[3] Lohse et al. (1997): Ermittlung von Verkehrsströmen mit n-linearen Gleichungssystemen – Verkehrsnachfragemodellierung. Schriftenreihe des Instituts für Verkehrsplanung und Straßenverkehr der TU Dresden.

[4] Behrens, R. (2000): Activity-Travel-Analysis: A review of theoretical origins, recent developments and local application. South African Transport Conference.

[5] Buliung, R.N., Kanaroglou, P.S. (2007): Activity-Travel Behaviour Research: Conceptual Issues, State of the Art, and Emerging Perspectives on Behavioural Analysis and Simulation Modeling. In: Transport Reviews, pp. 151-187.

[6] Bhat, C. et al. (2008): CEMDAP: Modelling and Microsimulation Frameworks, Software Development and Verification. In: Transportation Research Board Paper DVD, pp. 1-23.

[7] Pas, E. (1990): Is Travel Demand Analysis and Modelling in the Doldrums? In: Jones, P.: Developments in Dynamic and Activity-Based Approaches to Travel Anaylsis. Avebury.

[8] McNally, M.G. (2000): The activity-based approach. Working Paper. Institute of Transportation Studies at the University of California.

[9] Martínez, F., Justen, A., Cortés, C. (2009): A bi-level model of daily trip chains. International Time Use Observatory Workshop. Santiago, Chile.

[10] Bowman, J. L., Ben-Akiva, M. E. (2001): Activity-Based disaggregate travel demand model system with activity schedules. In: Transportation Research Part A, pp. 1-28.

[11] Beckman, R. J., Baggerly, K.A., McKay, M.D. (1996): Creating Synthetic Baseline Populations. In: Transportation Research Part A, pp. 415-429.

[12] Hobeika, A. (2005): Population Synthesizer. TRANSIMS Fundamentals, Chapter 3, pp. 1-117.

[13] Secretaría Interministerial de Planificación de Transporte – SECTRA (2002): Mobility Survey EOD 2001. Executive Report. Santiago, Chile.

[14] Zegras, C. (2004): Influence of Land Use on Transport Behaviour in Santiago, Chile. In: Transportation Research Record, pp. 175-182.