Accounting for similarities in destination choice modelling

A concept

Conference Paper

Author(s):
Schüssler, Nadine; Axhausen, Kay W.

Publication date:
2009-08

Permanent link:
https://doi.org/10.3929/ethz-a-005916843

Originally published in:
Arbeitsberichte Verkehrs- und Raumplanung 591

Accounting for similarities in destination choice modelling: A concept

Nadine Schüssler

Kay W. Axhausen

STRC 2009 **August 2009**

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

STRC 2009

Accounting for similarities in destination choice modelling: A concept

Nadine Schüssler
IVT
ETH Zürich
CH-8093 Zürich
phone: +41-44-633 30 85
fax: +41-44-633 10 57
schuessler@ivt.baug.ethz.ch

Kay W. Axhausen
IVT
ETH Zürich
CH-8093 Zürich
phone: +41-44-633 39 43
fax: +41-44-633 10 57
axhausen@ivt.baug.ethz.ch

August 2009

Abstract

Modelling the choice of destinations or activity locations is a crucial step in the analysis of transport behaviour. In real life, the decision makers' choices are highly complex and influenced by several factors such as the attributes of the destination itself, its accessability by different means of transport, the location of preceding or subsequent activities, or the attributes and accessability of competing destinations. In the traditional four step model, a lot of these factors have been ignored or only accounted for by rough approximations. State of the art transport models aim to overcome these shortcomings in different ways. Activity-based models incorporate trip chaining and accessability effects while classic discrete choice modelling focusses on capturing correlations between alternatives by developing more advanced model structures or similarity factors for individual attributes.

This paper presents a general framework for the treatment of similarity in a discrete choice model for destination choice of secondary activities. The framework combines several aspects of similarity derived from spatial location, the journey to and from the destination, trip chaining restrictions, and the attributes of the alternatives themselves. Moreover, it is applicable a simultaneous route, mode and destination choice model.

Keywords

Discrete choice, destination choice, similarities

1 Introduction

Modelling the choice of destinations or activity locations is a crucial step in the analysis of transport behaviour. In real life, the decision makers' choices are highly complex and influenced by several factors such as the attributes of the destination itself, its accessability by different means of transport, the location of preceding or subsequent activities, or the attributes and accessability of competing destinations. In the traditional four step model, a lot of these factors have been ignored or only accounted for by rough approximations. State of the art transport models aim to overcome these shortcomings in different ways. Activity-based models incorporate trip chaining and accessability effects while classic discrete choice modelling focusses on capturing correlations between alternatives by developing more advanced model structures or similarity factors for individual attributes.

This paper presents a concept to account for the different aspects of similarity between destination choice alternatives within the framework of discrete choice modelling. The focus is put on the modelling of *secondary* activities, i.e. activities for which the traveller can choose a new destination at the beginning of each trip. In contrast to that, the location of *primary* activities is determined in a long-term decision. Therefore, the location of primary activities should be modelled separately and can be treated as fixed in a model of daily travel behaviour. Classic primary activities are home, work or school, while shopping or meeting friends are typical secondary activities. The aim of this paper is to derive a general framework for the treatment of similarities in destination choice for secondary activities. The framework combines several aspects of similarity derived from spatial location, the journey to the destination, trip chaining restrictions, and the attributes of the alternatives themselves. Moreover, it is applicable a simultaneous route, mode and destination choice model.

The remainder of this paper is structured as follows. First, the basic modelling framework is introduced. Then, Section 3 outlines several concepts of how to integrate similarities in destination choice modelling. Subsequently, Section 4 presents the newly developed general framework for similarity treatment in destination choice modelling. The paper closes with conclusions and an outlook on the data requirements and future research issues regarding the modelling of destination choice and the treatment of similarities.

2 Modelling Framework

A destination choice model within the framework of discrete choice modelling is based on two assumptions. The first assumption is that the decision-maker, starting at a given origin, chooses one destination out of a set of available destination alternatives. The second assumption is that he or she uses utility maximisation as his or her decision rule. In the classic Multinomial Logit Model (MNL) model (McFadden, 1974), the utility U_{in} of alternative i for decision-maker n consists of two parts: a deterministic component V_{in}, defined by a vector β of taste coefficients and a vector x_{in} of attributes of the alternative, and a stochastic component ε_{in} comprising identically and independently (i.i.d.) Gumbel distributed error terms:

$$U_{in} = V_{in} + \varepsilon_{in} = f(\beta, x_{in}) + \varepsilon_{in} \tag{1}$$

The choice probability of alternative i can then be calculated as:

$$P(i|C_n) = \frac{e^{V_{in}}}{\sum_j e^{V_{jn}}} \tag{2}$$

An important aspect in the characterisation of destination choice alternatives is that they can be defined for different levels of spatial resolution. In modern micro-simulation models, destination choice alternatives consist of individual *facilities*, also called activity locations. A facility is defined as a synthetic entity where an activity can take place. Facilities represent, amongst others, factories, shops, schools, cinemas or apartment houses. They are described by a number of attributes such as location, opening hours or capacities. Since the number of facilities in an urban environment is huge, traditional destination choice models do not operate on facilities but on continuous spatial zones which contain several elemental facilities. In order to account for the aggregate nature of this type of destination choice alternatives Ben-Akiva and Lerman (1985) recommended the following adaption of the utility function:

$$U_{in} = \beta X_{in} + \frac{1}{\mu} ln M_i + \varepsilon_{in} \tag{3}$$

where M_i is the number of facilities in zone i and μ a scalar to be estimated. M_i should capture the combined attractiveness (expected utility) of all facilities in zone i often called *size of zone* i. μ represents common unobserved attributes of zone i that affect the attractiveness of its facilities. Ideally, μ would equal 1 implying that the model is invariant to configuration of spatial units. This paper, however, will deal with similarities in destination choice models on a more general level and derive a concept that is applicable to facilities and zones alike.

Regardless on the level of spatial resolution, the number of alternatives in a destination choice model is huge and manifold interactions between the alternatives are at work. The basic MNL model is not able to account for these similarities due to the prominent *Independence from Irrelevant Alternatives* (IIA) property. Independence from Irrelevant Alternatives means the relative ratio of the choice probabilities of two alternatives does not depend on the existence or the characteristics of other choice alternatives. The IIA property leads to biased parameter estimates and the disregard of an important aspect of the actual choice behaviour. Solving this issue is still an ongoing research topic as is the question whether similarities between alternatives have positive or negative effects on their choice probabilities. This question is linked to the mechanisms behind the behavioural reaction of the decision-maker when he or she faces similarities. Axhausen and Schüssler (2007) identified four main mechanisms:

- **Loosing visibility**: The alternative is undistinguishable from the other alternatives, leads to a lower probability to be included in the choice set and therefore to be chosen.
- **Joint risks**: Common elements lead, for the usually risk averse decision maker, to a reduction of the attractiveness of the alternative.
- Becoming a **super-alternative**: similar alternatives provide redundancy and therefore a higher chance of achieving one's goals. This increases the choice probability of all constituent alternatives.
- Gaining **super-visibility**: Being the best of a class of essentially similar alternatives increases the chances of being chosen both through more frequent inclusion into the choice set.

Since the effects of these four mechanisms can over-lay each other, similarities between destination choice alternatives can be very complex. Thus, suitable approaches have to accommodate various and complex similarity structures and at the same time be computationally efficient. In general, three ways to overcome the IIA property of the MNL have been pursued in the literature: allowing for non-zero elements in the variance-covariance matrix of the errors, factorial error components in addition to i.i.d. Gumbel errors and *adjustment terms* in the systematic part of the utility function. The approach of using adjustment terms is especially appealing because of its simplicity and elegance. Instead of structuring the choice set a priori or taking the chance of misleading assumptions about correlations, only the type of similarities is specified. That way, the individual characteristics of the alternatives are accounted for and a value is assigned to the impact of specific interdependencies. The most general formulation to include an adjustment term in the utility of an alternative can be found in Axhausen and Schüssler (2007):

$$U_{in} = f(\beta, x_{in}) + \alpha g(A_{in}) + \varepsilon_{in} \quad (4)$$

With $g(A_{in})$ being a transformation of the similarity term A_{in} and α a parameter to be estimated. Thereby, α can take positive as well as negative values because it is unclear a-priori how the similarity between alternatives will affect the choice behaviour.

Another approach that is computationally manageable for large sets of alternatives is the *Nested Logit model*. The basic idea of the Nested Logit model is to divide all alternatives of a choice set into disjoint nests. Correlations remain within a nest, but between the nests they are eliminated. Thus, the entire utility function for alternative i belonging to nest C_{nm} has to be reformulated. The systematic component is split into two parts and incorporates the alternative specific effects V_{in} as well as the impacts associated with the nest $V_{C_{mn}}$:

$$U_{in} = V_{in} + \varepsilon_{in} + V_{C_{mn}} + \varepsilon_{C_{mn}} \tag{5}$$

The distribution of the error-term ε_{in} remains IID Gumbel, while the error-terms $\varepsilon_{C_{mn}}$ jointly follow a generalised extreme-value distribution in a way that the random variable $max_{j \in C_{mn}} U_{jn}$ is Gumbel distributed with scale parameter μ. $V_{C_{mn}}$ is the composite utility of the nest C_{mn} , also called *expected maximum utility* or *Logsum*:

$$V_{C_{mn}} = V'_{C_{mn}} + \frac{1}{\mu_m} ln \sum_{j \in C_{mn}} e^{\mu_m V_{jn}} \tag{6}$$

where $V'_{C_{mn}}$ is the utility common to all alternatives in nest C_{mn}. Thus, the probability of choosing alternative i which is part of nest C_{mn} from the individual choice set C_n can be calculated as the product of the probability, that nest C_{mn} is chosen from the set of all nests and the probability that alternative i is chosen from the alternatives belonging to nest C_{nm}:

$$P(i|C_n) = P(C_{mn}|C_n) \cdot P(i|C_{mn}) \tag{7}$$

Since the Nested Logit model does not capture potential correlations between nests, it can only capture one aspect of similarity at a time. However, Vrtic (2003) demonstrated with the introduction of the Nested C-Logit (NCL) model that it can be extended straightforwardly by adding adjustment terms and remain computationally efficient.

Other approaches, e.g. the Cross Nested Logit model or the Mixed Logit model, are substantially more demanding in terms of computation time. They are not suitable for the problem of destination choice unless the analyst works with a very small sub-sample of the universal choice set and therefore neglected in the following. For a more in-depth discussion of these approaches the reader is referred to Schüssler and Axhausen (2007).

3 Account for similarities in destination choice modelling

Destination choice alternatives can be similar in manifold ways. This section gives now an overview about the different types of similarities and the approaches to account for these similarities by adding an adjustment term to the utility function. Along with adjustment terms that have already been established some new concepts will be introduced that have not yet been formalised but provide new ideas for adjustment terms. It is important to note that most of the concepts discussed in this section capture only one aspect of similarity. How these different concepts can be combined to a general framework for similarity treatment in destination choice modelling will be described in the next section.

3.1 Spatial proximity

The most obvious and most thoroughly discussed aspect of similarity is the spatial proximity between destination choice alternatives. One of the first researchers who accounted for spatial proximity in destination choice modelling was Fotheringham (1983) with his *competing destinations (CD)* model. The underlying assumption is a two step decision process: The decision-maker first chooses a broader region and second an alternative within that region. Therefore, the utility of each alternative is affected by the number of alternatives in the same region. With an increasing number of alternatives within the same region the probability for each alternative to be recognised, and thus to be chosen, decreases. Two formulations for the adjustment term have been presented so far: Fotheringham (1983) suggested to sum up the distances d_{ij} of a store i to all $I-1$ other stores j in the universal choice set and to weight the distances according to utility of the corresponding store.

$$CD_{in} = \left(\frac{1}{I-1} \sum_{j,j \neq i} \frac{V'_{in}}{d_{ij}} \right)^{\theta} \tag{8}$$

A second formulation has been proposed by Borgers and Timmermans (1987). It simply takes into account the average distance of store i to all other stores.

$$CD_{in} = \left(\frac{1}{I-1} \sum_{j,j \neq i} d_{ij} \right)^{\theta} \tag{9}$$

In both formulations θ is a parameter to be estimated. Following Fotheringham (1983)'s assumption that main similarity mechanism at work is the one of loosing visibility, the adjustment terms enter the utility function ln transformed and without an additional parameter that would allow for a positive impact of the similarity between alternatives.

According to Bernardin *et al.* (2009), the main weakness of the CD model is that it only measures the net effect of spatial proximity between destination choice alternatives while, in reality, there are two opposing forces at work: *spatial competition* and *agglomeration effects*. Spatial competition derives from similar alternatives located nearby and decrease the choice probability of an alternative because it looses visibility. The agglomeration effect, on the other hand, arises from alternatives that offer complementary goods or activities. The presence of complementary alternatives increases the choice probability since they facilitate trip chaining. Therefore, Bernardin *et al.* (2009) state that two adjustment terms have to be included in the utility function, each with its own parameter to be estimated. The two adjustment terms represent the accessibility of substitutes AS_i or complements AC_i from alternative i. In case the alternatives are facilities, the AS_i and AC_i can be derived straightforwardly. For the more complex case of zones, Bernardin *et al.* (2009) suggest the following formulations :

$$AC_i = ln \sum_j D_{ij} F_j e^{\alpha_C c_{ij}} \tag{10}$$

$$AS_i = ln \sum_j (2 - D_{ij}) F_j e^{\alpha_S c_{ij}} \tag{11}$$

where F_j is the total number of facilities in zone i, c_{ij} is the travel cost to get from zone i to zone j and α_C and α_S are parameters to be estimated. Even though α_C and α_S lead to an substantial increase in estimation time, Bernardin *et al.* (2009) prefer this accessibility formulation since the impact of the cost on the accessibility is determined by the data and not defined a-priori by the analyst. The most crucial variable, however, is the degree of dissimilarity D_{ij} between alternatives i and j. It is calculated by:

$$D_{ij} = 1 - \sum_c w_c \frac{F_{ic} F_{jc}}{F_i F_j} \tag{12}$$

where F_{ic} is the number of facilities of category c in zone i, F_i is the total number of facilities in zone i and w_c is a weighting function representing the number of times a facility of category c is visited in the study area. Bernardin *et al.* (2009) show that their model (slightly) outperforms the MNL and CD model and reacts behaviourally reasonable when a new alternative n is added to the choice set.

3.2 Trip Chaining effects

Closely related to the aspect of spatial proximity is the similarity derived from trip chaining. As Bernardin *et al.* (2009) discussed, the choice probability of an alternative can increase it is surrounded by complementary activities because the decision-maker is then able to execute several activities in one trip. Based on the same assumption, Kitamura (1984) developed a destination choice model that explicitly accounts for trip chaining effects by introducing an adjustment factor called *prospective utility* PU_{jn} which recursively integrates the utility that can be derived from subsequent activities in the utility of the destination under consideration:

$$PU_{in} = \sum_j q_{jn}(U_{jn}\theta d_{ij}) \tag{13}$$

where q_{jn} is the probability that decision-maker n carries out an activity at location j after his activity at location i,U_{jn} is the utility of said activity at location j, d_{ij} is the spatial distance between i and j, and θ is the disutility parameter for d_{ij}. PU_{in} can be interpreted as a measure of perceived accessibility of zone i. It can be modified to account for different trip purposes and due to it recursiveness also for longer trip chains.

A different approach that could be used to account for trip chaining effects is the *sequence alignment method (SAM)*. The SAM allows to determine the degree of similarity between entire trip and activity chains. It originates from molecular biology and was introduced into the field of travel behaviour research by Wilson (1998). Joh *et al.* (2001) and Wilson (2008) provided important enhancements. The SAM employs the concept of biological distance rather than geometrical distance. Biological distance is defined as the smallest number of attribute changes (mutations) that is necessary to equalise two sequences. Thus, the SAM allows to measure (dis)similarity regarding different attributes as well as the sequential order of activities. However, despite these advantages, no translation of the SAM dissimilarity measure into an adjustment term for discrete choice models has been presented so far.

3.3 Spatial learning

Another similarity aspect related to spatial proximity is spatial learning. Since a decision-maker can only make a choice between alternatives he knows, it is an ongoing research issue to determine how a decision-maker gets to know new destination choice alternatives. He or she might have been told about it by a friend or colleague, found it on the internet or, most importantly regarding the treatment of similarities, discovered it while travelling. One way to depict the spatial learning process and the resulting spatial knowledge of a decision-maker is to draw his or her *mental map*. Several studies (e.g. Chorus and Timmermans, 2009; Hannes *et al.*, 2008; Mondschein *et al.*, 2008) have been conducted to explore the relationships between mental maps on the one side and socio-demographics and travel behaviour on the other side. Their main findings are:

- Living in an area for a longer time improves the quality of the mental map.

- The regular use of modes that require active navigation (e.g. bike, car) improves the quality of the mental map significantly.
- Most activities have a standard mode-destination setting that is only changed if necessary.
- When deviating from their default option, decision-makers choose from a repertoire of standard alternative destinations that are often spatially linked to the default option.

This illustrates how much a destination choice set, and the choice probability of an alternative, depend on the places we have already visited. However, this issue has obtained little attention in the literature so far. One reason might be the lack of longitudinal survey data in the past. With the advent of more and more longitudinal diary and GPS studies, this obstacle should be overcome soon. As a first approach to account for effect of repeatedly visited destinations, Sivakumar and Bhat (2007) introduced a variable in the utility function which indicated whether the destination, in their case the zone, was chosen in the previous time period or not.

3.4 Travel mode related similarities

The travel modes we use do not only influence the set of destination alternatives we already know, but also determine which alternatives we consider accessible. Hannes *et al.* (2008) investigated the relationship between mode and destination choice. They showed, that mode and destination choice are usually executed simultaneously. Moreover, for each regularly performed activity people have a default mode-destination combination. Only if this default alternative is not available or suitable they choose a new mode-destination combination from a set of predefined alternatives. These standard sets are derived "from habit" or because the mode-destination combination is "logical". These findings are confirmed by Buliung *et al.* (2008), who examining location-based repetition in travel patterns. They concluded that people do not only conduct a big part of their activities at repeated locations but that they also visit these locations very often using the same transport mode. This effect is even stronger for public transport users than for car drivers.

Even though there is empirical evidence for the strong link between mode and destination choice, the transport mode still plays no significant role in destination choice modelling. Traditionally, the destination choice is carried out before the mode choice employing a combination of the travel costs of all modes to represent the accessibility by different transport modes. This does not do justice to the complexity of the relationship between mode and destination choice.

3.5 Dependencies between decision-makers

While the interaction between decision-makers is well acknowledged in route choice modelling and implemented by the means of assignment models, this aspect of similarity has been neglected in destination choice models so far. One suitable way to integrate dependencies between decision-makers is the *field effect* approach by Dugundji and Walker (2005). This adjustment term represents the share of other, connected decision-makers who choose the same alternative. The connections between the decision-makers can be caused by different factors and are formalised by the means of a network. This network is then used to calculate the field effect variable.

A different approach, that was not integrated in a discrete choice model but an agent-based simulation, was presented by Horni *et al.* (forthcoming). In their model, the utility of an activity depends on the ratio between the load and the capacity of a facility. The introduction of this term inducing a penalty for overcrowded facilities, which influences the agent's activity location choice and leads to more realistic results.

3.6 Attribute derived similarity

The next type of similarity derives from the attributes of the destination choice alternatives themselves. Different measures to determine the attribute similarities have presented in the literature. Amongst them are the similarity measures of Batsell (1982)

$$SIM_i = exp(\frac{1}{I-1}\sum_j \sum_k \theta|x_{ik} - x_{jk}|), \tag{14}$$

Borgers and Timmermans (1987)

$$SIM_i = \prod_k \left[\frac{1}{I-1}\sum_j |x_{ik} - x_{jk}|\right]^{\theta/K}, \tag{15}$$

and Meyer and Eagle (1982)

$$SIM_i = \left[\frac{1}{I-1}\sum_{j\neq i} 0.5|r_{ij} - 1|\right]^{\theta}. \tag{16}$$

where x_{ik} is the value of attribute k for alternative i, r_{ij} is the observed Pearson product moment correlation between alternatives i and j across their attributes, θ a parameter to be estimated and I the total number of alternatives.

A more sophisticated measure is the *concept of dominance* introduced by Cascetta and Papola

(2005). It is based on the assumption that an alternative is less likely to be taken into account if it is dominated by other alternatives. Alternative j dominates alternative i, if the utility of all attributes of j is higher than (or equal to) the utility of the equivalent attributes of i. Following that concept, a dominance factor DF_{in} is calculated for each alternative i, indicating the number of alternatives dominating i. Cascetta and Papola (2005) define two specifications for the dominance measures. In the first specification, they assume that alternative j dominates alternative i if the utility of j is greater than that of i for all attributes of i and j while at the same time the generalised costs c_{oj} of getting from origin o to destination j are smaller than those of getting from o to i. The second dominance measure originates from the concept of intervening opportunities (Stouffer, 1940). In order to dominate i, destination j has to fulfil the conditions formulated above and, in addition, has to be situated on the path from origin o to destination i. In this case j is an intervening opportunity on the path to i.

Martínez *et al.* (2008) used the dominance factor as cut-off value for their Constrained Multinomial Logit Model which models the probability of an alternative to be included in the individual choice set of the decision-maker with a binomial logit function. They detected that the dominance affects the utility in a non-linear way. Accordingly, further research is advised concerning the way the dominance factor should enter the utility function. Non-linear transformations should be tested as well as minimum or maximum thresholds.

3.7 Image of a destination

While most of the similarity aspects discussed above can be measured on a continuous or ordinal scala, the last aspect of similarity is purely nominal. Huybers (2003) and Hong *et al.* (2006) demonstrated how strongly the image of a destination can influence the choice behaviour. Huybers (2003) investigated the difference in choice behaviour for labeled and unlabeled vacation destinations and observed that the choice behaviour and the resulting market shares differed considerably. Evidently, the labels of the vacation destination carried a lot more meaning than the attributes considered in the survey. This meaning is often referred to as the *image of a destination*. Hong *et al.* (2006) experienced similar effects when they examined the effect of categorisation and the image of these categories on vacation destination choice. They derived four categories (mountainous, coastal, historic and exotic) and proved by the use of a Nested Logit model, that the alternatives within these categories are perceived to be more similar to each other than to destinations of other categories and that this similarity influenced the choice decision. Analogously, a supermarket chain can carry a certain image that goes beyond the range of goods offered and the overall price level and influences people's shopping destination choice. Thus, the image of a destination is an important aspect of similarity and should be accounted for in a destination choice model.

4 A general framework of similarity treatment in destination choice modelling

The concepts to account for similarities in destination choice, that have been described in the last section do only capture one aspect of similarity at a time even though the different aspects interact with each other. The travel mode, spatial proximity, trip chaining and spatial learning do all influence each other whereas the image of an alternative depends on the attributes of that alternatives as well as on the behaviour of other decision makers. In order to derive a comprehensive understanding of similarities between destination choice alternatives the different aspects and their interactions have to be evaluated and weighted against each other. This section does this on a conceptual level. The empirical confirmation is a future research issue.

Several assumptions about the destination choice model itself are necessary to derive a framework for similarity treatment: The complete daily schedule including the order and type of activities, their timing and durations, is known. The location of the primary activities is fixed and known, as is the equipment of the household with mobility tools such as cars or public transport subscriptions. The universal choice set for each activity comprises all facilities or zones in the study area where an activity of the specified type can be carried out. Since the universal choice set is too large for a destination choice model, it has to be reduced appropriately to the individual choice set. A first reduction is done using the shortest path-based space-time prism approach developed by Scott (2006). The space-time prism approach is better suited than the traditional way of including all alternatives within a circle around the origin because it accounts for the impact of the subsequent primary activity. Additionally, it is able to deal with sequences of secondary activities between two primary activities (i.e trip chaining effects) and can derive different choice sets for different modes. Since the number of alternatives within a space-time prism is still very large, it is advised to use only a sub-sample of the alternatives and the version of the algorithm presented by Horni *et al.* (forthcoming) with improved computational performance for this kind of task. How to draw the sub-sample is an important question but goes beyond the scope of this paper. In a basic version, the alternatives could be drawn randomly. However, other sampling procedures, more tailored to the requirements of destination choice sets, should be tested as well. Finally, the choice set generation procedure has to ensure that all alternatives within the space-time prism that have been chosen by the decision-maker for the activity type in question during the survey period are also part of the choice set.

Given these prerequisites, the following aspects of similarity will be accounted for in the framework described in this section:

- similarity derived from travel mode
- similarity caused by spatial proximity

- similarities emerging from spatial learning and spatial repetition, and
- similarities originating from the image of the destinations.

Trip chaining effects will not be considered explicitly because they are already captured to some extent by the space-time prisms approach in the choice set generation and to another extent in the spatial proximity adjustment terms. Dependencies between decision-makers are also neglected. They play either a minor role, e.g. for shopping trips, or are so important that an explicit modelling, e.g. through social network approaches, would be necessary.

The **travel mode** has a strong influence on the choice set, because the travel cost to the same location can differ significantly depending on the mode. Some alternatives might even become inaccessible. In addition, some modes might not be available for all decision-makers. Therefore, it is very important to account for mode similarities in destination choice. The proper way to do this would be a combined mode and destination choice model. This could be best described in a Nested Logit model. The modes available to the decision-maker constitute the nests. The availability of a mode can be based on the household's mobility tools and further information derived for example from the daily schedule. The destinations in each nest should then be generated separately for each mode. Another option would be to derive a common choice set for all modes and to introduce a similarity term based on mode accessibility. However, this will hardly be able to depict the interactions between mode and destination choice as precisely as a combined mode and destination choice model.

Following the argument of Bernardin *et al.* (2009), the effects of **spatial proximity** of substitutes and complements should be accounted for separately. But a different measure for spatial proximity is proposed here. A buffer is created around the destination and all (complementary or substitutional) alternatives within that buffer are summed up weighted by the distance from the destination in question using a distance decay function. The form of the distance decay function and the size of the buffer should be derived from empirical data and are future research issues. Furthermore, not only alternatives around the destination influence the choice of the decision-maker but also alternatives on the way to and from the destination. Thus, these alternatives are also incorporated in the measure of spatial proximity employing the same concept of a buffer and a weighting by distance. Thereby, the construction of the buffer along the way has to follow the characteristics of the modes. For public transport trips, it suffices to create a buffer around the public transport stops, whereas for private transport modes a buffer around the complete path is suitable. This, however, requires a combined route, mode and destination choice model. If a combined route, mode and destination choice model is not applicable, for example because of the lack of data on the chosen routes, a possible work around was provided by Cascetta and Papola (2005) with their determination of alternatives "along the path". In their definition, an alternative k is located along the path from o to d if

$$dist(o,k) + dist(k,d) < 1.2 \cdot dist(o,d). \tag{17}$$

Moreover, a distinction has to be made between models in that work on the spatial resolution of travel zones and those in which facilities compose the decision entities. In case of facilities as alternatives, the spatial proximity effects arising from complements SPC_i and substitutes SPS_i for alternative i should be accounted for by:

$$SPC_i = \sum_{j \epsilon Co} w(d(i,j)) \cdot \delta_j \tag{18}$$

$$SPS_i = \sum_{j \epsilon S} w(d(i,j)) \cdot \delta_j \tag{19}$$

where Co is the set of alternatives that are complements to the activity in question, S is the set of alternatives that are substitutes for the activity in question, $w(d(i,j))$ is the distance decay function used as weight for alternative j and δ_j is an indicator function which equals 1 if j lies within the overall buffer area and 0 otherwise.

If the alternatives of the destination model are zones, it is necessary to determine for each zone the "size" of the complementary and substitutional facilities within each zone. The size attribute can be measured in various ways, e.g. based on the number of facilities, the number of employers in these facilities or the square footage of the facilities. Then, the size variables sc_j and ss_j have to be incorporated in the spatial proximity measures SPC_i and SPS_i. Moreover, the distance $d(i,j)$ between alternative i and alternative j is measured as distance between the zonal centroids. Finally, it it has to be defined when a zone lies within the buffer, i.e. when δ_j obtains a value of 1. Several variants are possible: The buffer has to cover the zone completely, a certain percentage of the zone or the at least the centroid. Another option would be to transform δ_j into a continuous variable ranging from 0 to 1 and representing the percentage of the zone that lies within the buffer. Merging all these considerations, the spatial proximity measures for zone alternative i are:

$$SPC_i = \sum_{j \epsilon C} w(d(i,j)) \cdot sc_j \cdot \delta_j \tag{20}$$

$$SPS_i = \sum_{j \epsilon S} w(d(i,j)) \cdot ss_j \cdot \delta_j \tag{21}$$

In order to measure the similarity effects originating from spatial learning or **spatial repetition**, multiple-day observations for each survey participant are required. Following the approach suggested by Sivakumar and Bhat (2007) a similarity measure based on spatial repetition is suggested. In case of zones, the spatial repetition index should indicate the time period, e.g. the number of days, since the zone was last chosen. The spatial repetition measure SR_i is then formulated as:

$$SR_i = \frac{t_0 - t_i}{t_0} \tag{22}$$

where t_0 is the current period of time and t_i is the period of time when the zone i was last visited. Both times should be measured relative to the beginning of the survey period. This way, the beginning of the survey period can be defined as the default value for each t_i. Together with the normalisation by the current time this leads to a range between 0 and 1 for SR_i. 0 indicates that the zone was already visited in the same period of time and 1 that it was never visited before in the survey period.

In a model where the alternatives are elemental facilities, the definition of repetition is extended in the sense that not only a visit to the same facility is counted but also to other facilities that are close by. This can be done by either considering all facilities within a circle around the destination in question or by summarising alternatives to neighbourhoods and determining the last time the person visited the neighbourhood. The spatial repetition measure SR_i is then formulated analogously to the one for the zones:

$$SR_i = min_{j \epsilon N} \frac{t_0 - t_j}{t_0} \tag{23}$$

where N is the set of facilities that lie within the circle or neighbourhood around i, including i. For t_0 and t_j the same definitions as above apply.

The derivation of the **image of an alternative** strongly depends on the choice context, the level of spatial resolution and the characteristics of the alternatives that are available. In grocery shopping, for example, the image of the supermarket can be based on the chain it belongs to, the size of the shop, the range of goods or on its crowdedness whereas vacation destination can build up their image through their location, the price level or special attractions. Therefore, the analyst first has to decide which characteristics are crucial for the image of a destination. Based on this, the alternatives are subdivided into categories. The categorisation can enter the utility function either by the means of dummy variables and/or in the form of a measure for image related similarity IM_i:

$$IM_i = \frac{\sum_{j \epsilon C, j \neq i} \delta_{jy}}{I - 1} \tag{24}$$

where C is the complete choice set, I is the number of alternative in C, and δ_{jy} is an indicator function that takes the value of 1 if alternative j belongs to the same category y as alternative

i. The values of the similarity measure range between 0 and 1, zero indicating that no other alternatives belongs to the same category and 1 that all alternatives belong to the same category.

Finally, the four similarity components have to be combined to one model that accounts for all the different aspects of similarity. As described above the model is a Nested Logit model for combined mode and destination choice. Thus, the utility function bases on Equation 5. As in Vrtic (2003), the adjustment terms are added to deterministic parts of the utility of each destination alternative. Thereby, each adjustment term obtains it's own parameter α_l. This parameter is estimated and can take positive as well as negative values because it is unclear a-priori how the similarity between the alternatives will affect the choice behaviour. The utility U_{in} of an destination alternative i, mode m and decision-maker n is then formulated as:

$$U_{in} = V_{in} + \varepsilon_{in} + V_{C_{mn}} + \varepsilon_{C_{mn}} \tag{25}$$

where ε_{in} and $\varepsilon_{C_{mn}}$ are the error terms of the Nested Logit model following the distributional assumptions described in Section 2. The deterministic part of the utility V_{in} of each destination alternative is described by:

$$V_{in} = f(\beta, x_{in}) + \alpha_1 SPC_i + \alpha_2 SPS_i + \alpha_3 SR_i + \alpha_4 IM_i, \tag{26}$$

while the Logsum term for each mode is calculated by:

$$V_{C_{mn}} = V'_{C_{mn}} + \frac{1}{\mu_m} ln \sum_{j \in C_{mn}} exp(\mu_m(f(\beta, x_{in}) + \alpha_1 SPC_j + \alpha_2 SPS_j + \alpha_3 SR_i + \alpha_4 IM_i)), \tag{27}$$

where $V'_{C_{mn}}$ is the utility common to all alternatives in nest C_{mn}.

5 Conclusion and outlook

Destination choice situations are characterised by a multitude of similarities originating from different sources. This paper offers a framework that enables the analyst to account for several similarity aspects simultaneously. The framework is build in a modular way to make it adjustable in case the required data is not available. Moreover, for each similarity component, a separate coefficient has to be estimated to depict the relative influence of the similarity component on the overall choice.

No specific similarity treatment is proposed for trip chaining effects. Instead it is advised to use a state of the art choice set generation procedure, the space-time prism approach, which by design takes care of the impact of trip chaining on the choice set. Since destination choice sets strongly depend on the mode and behavioural research demonstrated that people choose mode and destination simultaneously, a combined mode and destination choice model is suggested. The combined model takes the form of a Nested Logit model where the modes are the nests and the destinations are the alternatives within the nests. The spatial proximity adjustment term does not only differentiate between complementary and substitutional alternatives, it incorporates, in addition to the alternatives around the destination in question, those on the way to and from the destination. Ideally, this is done in a combined route, mode and destination choice model. A work around is presented in case such a model is not applicable. Another important aspect of similarity is spatial repetition and spatial learning. The adjustment term proposed in this paper accounts for the time that has passed since the destination was visited last. Finally, the similarities in the image of alternatives do also influence the destination choice. The derivation of the image attribute, however, depends on the choice context and the available data.

In order to estimate models based on the framework proposed in this paper, some data requirements have to be met. Most importantly, the survey data has to have the form of a diary which records the complete daily schedule of the participant, including times of journeys and activities, the types of activities and the mode used for travelling. Additionally, multi-day records would be necessary to include similarity based on spatial repetition and observations aboute the routes for a combined route, mode and destination choice model. Concerning the infrastructure, the data should be as detailed as possible. Even though the framework is applicable to zonal data, it is primarily designed to account for similarities between elemental facilities.

Another important issue, that has not been discussed in this paper because it is only loosely related to the question of similarities, is the way the travel time is considered in destination choice models. While an absolute continuous value for the travel time might be appropriate for route choice models it is arguable for destination choice models. Travel time is more likely to be perceived in intervals and relative to the travel time to other destinations. Thus, an important research issue is to investigate different ways of including travel time in the destination model.

As pointed out in section 4, there are also several open research issues concerning the framework for similarity treatment. Regarding the choice set generation the main question is how a sub-sample out of all the alternatives within the space-time prism should be drawn. With respect to the spatial proximity measure the correct size of the buffer, the form of the distance decay function, the derivation of the size variables sc_j and ss_j and the definition of which zones lie within the buffer area have to be investigated. In addition, a more detailed literature research which exceeds the field of transportation research, might help to determine how the image of an alternative is formed and how it can be derived from available data.

6 Acknowledgements

The authors would like to thank the Swiss National Science Foundation for the funding of this work.

References

Axhausen, K. W. and N. Schüssler (2007) Similarity, visibility and similar concepts: A note, *Working Paper*, **458**, IVT, ETH Zurich, Zurich.

Batsell, R. R. (1982) A multiattribute extension of the luce model which simultaneously scales utility and substitutability, in J. Huber (ed.) *The Effect of Item Similarity on Choice Probabilities: A Collection of Working Papers from a Conference at Quail Roost*, 4–23, Duke University, Durham.

Ben-Akiva, M. E. and S. R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge.

Bernardin, V. L., F. S. Koppelman and D. Boyce (2009) Enhanced destination choice models incorporating agglomeration related to trip chaining while controlling for spatial competition, paper presented at *the 88th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2009.

Borgers, A. W. J. and H. J. P. Timmermans (1987) Choice model specification, substitution and spatial structure effects: A simulation experiment, *Regional Science and Urban Economics*, **17** (1) 29–47.

Buliung, R. N., M. J. Roorda and T. K. Remmel (2008) Exploring spatial variety in pattterns of activity-travel behaviour: Initial results from the Toronto Travel-Activity Panel Survey (TTAPS), *Transportation*, **35** (6) 697–722.

Cascetta, E. and A. Papola (2005) Dominance among alternatives in random utility models: A general framework and an application to destination choice, paper presented at *European Transport Conference*, Strasbourg, October 2005.

Chorus, C. G. and H. J. P. Timmermans (2009) An empirical study into the influence of travel behavior on stated and revealed mental maps, paper presented at *the 88th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2009.

Dugundji, E. R. and J. L. Walker (2005) Discrete choice with social and spatial network interdependencies: An empirical example using mixed gev models with field and "panel" effects, paper presented at *the 84th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2005.

Fotheringham, A. S. (1983) A new set of spatial interaction models: The theory of competing destinations, *Environment and Planning A*, **15** (1) 15–36.

Hannes, E., D. Janssens and G. Wets (2008) Does space matter? travel mode scripts in daily activity travel, *Transportation Research Record*, **2054**, 20–27.

Hong, S. P., J.-h. Kim, H. Jang and S. Lee (2006) The roles of categorization, affective image and constraints on destination choice: An application of the NMNL model, *Tourism Management*, **27** (5) 750–761.

Horni, A., D. M. Scott, M. Balmer and K. W. Axhausen (forthcoming) Location choice modeling for shopping and leisure activities with MATSim: Combining micro-simulation and time geography, *Transportation Research Record*.

Huybers, T. (2003) Destination choice modelling - to label or not to label?, paper presented at *the Tourism Modelling and Competitiveness Conference*, Paphos, Cyprus, November 2003.

Joh, C.-H., T. A. Arentze and H. J. P. Timmermans (2001) A position-sensitive sequence alignment method illustrated for space-time activity-diary data, *Environment and Planning A*, **33** (2) 313–338.

Kitamura, R. (1984) Incorporating trip chaining into analysis of destination choice, *Transportation Research Part B: Methodological*, **18** (1) 67–81.

Martínez, F., E. Cascetta, F. Pagliara, M. Bierlaire and K. W. Axhausen (2008) An application of the Constrained Multinomial Logit (CMNL) for modelling dominated choice alternatives, paper presented at *the 8th Swiss Transport Research Conference*, Ascona, October 2008.

McFadden, D. (1974) Conditional logit analysis of qualitative choice-behaviour, in P. Zarembka (ed.) *Frontiers in Econometrics*, 105–142, Academic Press, New York.

Meyer, R. J. and T. C. Eagle (1982) Context-induced parameter instability in a disaggregate-stochastic model of store choice, *Journal of Marketing Research*, **19** (1) 62–71.

Mondschein, A., E. Blumenberg and B. D. Taylor (2008) Accessibility and cognition: Effect of transportation mode on spatial knowledge, paper presented at *the 87th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2008.

Schüssler, N. and K. W. Axhausen (2007) Recent developments regarding similarities in transport modelling, paper presented at *the 7th Swiss Transport Research Conference*, Ascona, September 2007.

Scott, D. M. (2006) Constrained destination choice set generation: A comparison of GIS-based approaches, paper presented at *the 85th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2006.

Sivakumar, A. and C. R. Bhat (2007) Comprehensive, unified framework for analyzing spatial location choice, *Transportation Research Record*, **2003**, 103–111.

Stouffer, S. A. (1940) Intervening opportunities: A theory relating mobility and distance, *American Sociological Review*, **5** (6) 845–867.

Vrtic, M. (2003) Simultanes Routen- und Verkehrsmittelwahlmodell, Ph.D. Thesis, Technical University Dresden, Dresden.

Wilson, C. (1998) Analysis of travel behavior using sequence alignment methods, *Transportation Research Record*, **1645**, 52–59.

Wilson, C. (2008) Activity patterns in space and time: Calculating representative Hagerstrand trajectories, *Transportation*, **35** (4) 485–499.