

# Choice Set Generation for Multi-modal Travel Analysis

Stella Fiorenzo-Catalano, Rob van Nes and Piet H.L Bovy  
Transport and Planning Section  
Faculty of Civil Engineering and Geosciences  
Delft University of Technology  
Delft  
The Netherlands  
E-mail: S.Catalano@ct.tudelft.nl

EJTIR, 4, 2, (2004), pp. 195-209

*Received: February 2004*

*Accepted: May 2004*

*Multi-modal trips are a common travel phenomenon, which are expected to become more important in the future because of their expected contribution to sustainable urban transportation. However, multiple different types of travel choices, such as transport service types, travel modes, and transfer locations, are involved in a multi-modal trip, making it difficult to model multi-modal traveling. We present a method that generates choice sets of multi-modal routes using a supernetwork, which might be used for prediction purposes. This method considers stochasticity in the perception of the network attributes as well as in the preferences for the different trip components. The primary objective of the paper is to analyze the comparison of generated route sets and observed route sets. Three options for generating route sets have been studied, i.e.: variation in the network attributes only, variation in traveler preferences only, and the combination of both. The latter case proved to yield the best match with observed route sets. Furthermore, the analysis shows that variation in travelers' preferences is more important than variation in network attributes. Recommendations for further improvement of the choice set generation method are included. The analysis revealed insights into the possibilities of generating realistic multi-modal route sets and it is proved that the randomization approach is feasible providing good coverage of the observed routes. By far the best results are obtained by randomizing both network attributes and variation in traveler preferences.*

**Keywords:** Choice sets, Multi-modal, Public transport

## 1. Introduction

Multi-modal trips, i.e. trips consisting of two or more vehicular modes, are a common travel phenomenon, which are expected to become more important in the future. Although multi-modal trips overall only account for less than 3% of all total passenger transport in The Netherlands, it is noted in Van Nes (2002) that multi-modal transport merits attention as it serves an important market within urban and inter-urban transport. For instance, over 20% of the inter-urban trips to and from the larger Dutch cities are multi-modal trips with usually train as the main transport mode.

The specific theoretical challenge in modeling multi-modal trips is in the multi-dimensional character of these trips. Multi-modal passenger transport modeling must deal with the simultaneous choice of routes, travel modes, and interchange locations between public transport modes, access/egress locations from private to public transport modes and vice versa.

In order to analyze this complex topic we adopt a route-based approach, e.g. Bovy and Stern (1990) and Fernandez *et al.* (1994), i.e. we assume that a traveler has a set of possible multi-modal route alternatives available for a specific trip (i.e. choice set) from which (s)he chooses the alternative that is most suited for his travel need. In this paper a route alternative is defined as a sequence of modes and intermediate transfer nodes the traveler uses to make a trip from an origin to a destination in the multi-modal network. In this context a mode is defined as a transport service type in a vehicular or functional sense.

Route choice set enumeration consists in finding all feasible routes that a traveler might consider for traveling from his origin to his destination. A priori enumeration in a network context not only offers a number of theoretical advantages related to travel choice modeling such as inclusion of non-linear cost functions and route-specific attributes; it also offers implementation and computational advantages in iterative network assignment approaches since no repeated optimal route search is necessary. This has, for instance, been demonstrated in dynamic equilibrium modeling of a large road network (Bliemer *et al.*, 2003 and Bliemer *et al.*, 2004).

In transportation modeling it is not common to follow a completely individual-level approach, therefore we have to distinguish between individual and group level. Especially in predicting choice behavior the analyst often has to resort to a group-level approach. If instead of a single traveler, a set of travelers having similar demand conditions, traveling between similar origin and destination areas, and having similar preferences and characteristics is considered, the choice set consists of all multi-modal route alternatives available to the set of travelers that satisfy their travel needs.

In this paper we will focus on route choice set generation for multi-modal trips: that is generating a set of realistic multi-modal routes (modes, service types, transfer nodes, etc.) that a group of travelers might consider for making a multi-modal trip. A route set generation method is presented which explicitly considers stochasticity in different components, i.e. network attributes and travelers' preferences (see e.g. Nielsen, 1996, Nielsen 2002, and Ramming, 2002). This method is applied to a realistic multi-modal supernetwork while considering the stochasticity of either of these components as well as of the combination of the two. The primary objective of the paper is to compare the generated choice sets with observed multi-modal route sets from individual travelers. This analysis provides insights into the possibilities of generating realistic multi-modal route sets, as well as insight into the

consequences of considering the stochastic nature of both the network and travelers' behavior.

The paper is structured as follows. In section 2 some theoretical concepts with respect to choice sets are introduced. Section 3 gives a short overview of methods for generating choice sets and describes the proposed stochastic generation approaches. Section 4 deals with the application of three approaches in a case study. In section 5 a comparison between the routes sets generated by each of the three approaches and a sample of observed route sets is discussed. Finally, section 6 presents some conclusions on observed and generated choice sets and recommendations for further research.

## 2. Theoretical notions and terminology

An overview of the elements of individual route choice behavior is given in Bovy and Stern (1990), with emphasis on the way route finding is structured. The key notion is that travelers consider a set of routes for a trip. The concept of route sets poses several interesting problems. For instance, on the one hand, the number of alternatives for a specific Origin-Destination-(OD)-relation may be large, especially in urban road networks and in multi-modal networks. On the other hand, individual travelers only know a subset of all existing feasible route alternatives while having a certain perception of the route characteristics following from their travel experience and information acquisition behavior. Consequently, the number of alternatives considered by an individual traveler will be substantially smaller. In addition, traveler's knowledge of the transport network, his so-called *mental map*, strongly depends on how the objective network and route attributes are perceived and distorted into the subjective route factors that are successively evaluated for the decision process. Furthermore, there is the difference between an individual and a group of travelers. While individual consideration sets might be small and may have different compositions, the union of these sets over all travelers between an OD-pair might be considerable (Bovy and Stern 1990).

By definition, the *choice set* is the set of all route alternatives available to a traveler, constituted by the subset of known alternatives satisfying the travel need of the individual. Two main types of choice sets can be distinguished: the *subjective* choice set, consisting of the trip alternatives known by the individual traveler, and the *objective* choice set, consisting of all feasible alternatives considered relevant by the researcher for the traveler. Furthermore, objective choice sets might also be defined for populations of travelers having the same OD-relation.

The approach presented in this paper is developed for generating realistic route sets for groups of travelers. Since the objective choice sets might be very large (Hoogendoorn-Lanser and Van Nes, 2004), the method aims at generating a realistic subset. Such a route set should at least contain the subjective choice set, or in the case of a population of travelers, the union of subjective choice sets, while no wrong or unrealistic routes should be generated.

### 3. Methods for generating route choice sets

Route sets might be generated for a variety of purposes, such as estimation, prediction, and data enhancement. Literature shows a large variety of techniques for generating routes, for a recent account see Ramming (2002). Typical approaches are the K-shortest path algorithm (Van Der Zijpp and Fiorenzo-Catalano, 2002) based on link elimination or link penalties (De La Barra et al., 1993), simulation methods (Nielsen, 1996, Nielsen, 2002 and Sheffi and Powell, 1982), and a labeling approach (Ben-Akiva et al., 1984). Available route choice literature appears to credit simulation approaches a dominant performance (Ramming, 2002, and Fiorenzo-Catalano and Van Der Zijpp, 2001).

For this study a combination of the labeling method and the simulation method is used and applied to a so-called supernetwork (Carlier *et al.*, 2003). The supernetwork consists of the concatenated networks of all modes, i.e. walking, cycling, car driver, car passenger, urban and interurban public transport services, and of 'boarding' and 'alighting' links between each single mode network and the walk-network. The latter links enable travelers to switch modes during a trip. The public transport service network is represented using lines and frequencies. Since in this method no use is made of timetable data, the proposed choice set generation approach is valuable for static (frequency-based) trip assignment. However, it can be easily adapted for dynamic (scheduled-based) assignments as well (Nuzzolo *et al.*, 2001)

The labeling method distinguishes different groups of travelers for whom the most attractive paths are determined in a multi-modal supernetwork using randomized generalized costs and a shortest path algorithm. Selecting alternative routes by adopting a shortest path procedure reflects on the theoretical point-of-departure that individuals try to choose the route that minimizes their subjective disutility. Furthermore, the use of the shortest path criterion eliminates the chance that wrong routes are generated, that is, if realistic parameters for the generalized costs are used.

The adopted generalized cost function synthesizes the most important multi-modal trip attributes and their weights as known from earlier studies (Van Der Waard, 1988, Nielsen, 1996, and Wardman, 2001). In principle, generalized cost functions will differ by user class or trip purpose  $s$ . Please note that in the supernetwork every link relates to a single mode and a single trip component. For our generation purpose we assume that the travel cost  $c$  of the multi-modal path  $p$  is a summation of link costs (neglecting route-specific non-link costs or non-linear cost structures):

$$c_p^s = \sum_{\forall a \in p} c_a^s \quad (1)$$

where  $c_a^s$  is the travel cost  $c$  on link  $a$  for user class  $s$ . Link cost  $c_a^s$  is considered to be a stochastic quantity reflecting stochastic variation of attribute perceptions and attribute preferences among travelers.

The formulation of the link cost function is as follows:

$$c_a^s = \alpha_m^s * C_a + VOT^s * \beta_m^s * X_a + CK_m^s * D_a \quad (2)$$

where  $C_a$  is a link specific cost, which may represent e.g. toll cost for road links or parking cost for parking links.  $X_a$  is the time attribute of link  $a$ , may be travel time or waiting time depending on the link type while  $D_a$  is the length of link  $a$ .  $VOT^s$  is value of time for user

class  $s$  while  $CK_m^s$  is the cost per kilometer for user class  $s$  and mode  $m$ .  $\alpha_m^s$  and  $\beta_m^s$  are the weights of perceived attributes based on transport mode  $m$  and user class  $s$ . The parameters used in the link cost function ( $\alpha_m^s$  and  $\beta_m^s$ ) are specific for each attribute (such as travel time, waiting time, etc.) and depend on personal preferences of the users, although in most cost models it is assumed that  $\alpha=\beta=1$ .

This travel cost approach reflects our hypothesis that the composition of individual choice sets is strongly determined by individual preferences for trip attributes. It is well known (see e.g. Ben-Akiva and Bierlaire, 1999) that travel time is the most important attribute to be included in the utility function, although additional link attributes such as travel costs and distance should also be considered. In our model we have taken into account link attributes that are usual for private and public transport trip parts (see Ben-Akiva and Bierlaire, 1999) supplemented for attributes that are typical for the transfer movement between modes. In particular, length and travel time are taken into account as attributes for private modes (car and bike), whereas in-vehicle travel time is taken into account for public modes (train, bus, tram and metro). In addition to the travel link attributes, the following attributes associated with transfer links are considered: waiting time, boarding time, alighting time, parking time and parking cost (last one only for car).

In the combined labeling and randomization procedure we distinguish traveler groups that vary with respect to expected travel behavior, for instance based on trip purpose, and vehicle availability at the home-end and the activity-end of the trip. For the randomization we consider two options. Given that travel times and other time components along different routes may vary from day-to-day due to service fluctuations, traffic lights, congestion, weather condition, etc. one approach to generate route sets is to consider stochasticity at the network level. This also includes attribute perception errors from the travelers' perspective. In this case, the link attributes ( $X$ ) are randomized using Monte-Carlo techniques (Sheffi and Powell, 1982) weighted according to the average preferences of the specific traveler groups (or user classes). Secondly, individual travelers may have different preferences for trip attributes as well and may choose based on personal perceptions. These preferences are described by the weights used in the utility function (behavioral parameters:  $\alpha$ ,  $\beta$ ,  $VOT$ , and  $CK$ ), which may be randomized as well to reflect the variability in observed travel behavior (Nielsen, 1996 and Nielsen, 2002). For practical reasons, only the  $\beta$  parameters are randomized in the presented analyses, while the others ( $\alpha$ ,  $VOT$ , and  $CK$ ) are kept constant.

Scheme 1 shows an algorithmic outline of the proposed approach to generate subjective choice sets. First, a randomized network is generated by sampling the link attributes ( $X$ ) from some positive statistical distribution (e.g. truncated normal or Gamma distribution). Next, for each traveler group the parameters of the link cost function ( $\beta$ ) are sampled, again from some (positive) statistical distribution, followed by computing the generalized link costs with respect to randomized link attributes and parameters. Then, for each OD pair the minimum-cost path with respect to the generalized route cost is computed and the new path is inserted in the path list, that is, if the route hasn't been found yet. To generate a sufficiently high number of routes and to achieve sufficient variation of the routes the process of sampling link attributes and parameters is repeated for a given number of iterations (steps  $a$  and  $d$  respectively).

- (a) Repeat for a given number of iterations
- (b) Sample the link attributes ( $X$ ) from a positive statistical distribution
- (c) For each traveler group
  - (d) Repeat for a given number of iterations
  - (e) Sample the parameter values ( $\beta$ ) of the link cost function from a positive statistical distribution
  - (f) Compute generalized link cost with respect to randomized both link attributes  $X$  and  $\beta$  parameters
  - (g) For each OD pair search for the shortest path based on the randomized generalized link costs
  - (h) Insert the new path in the path list if it has not been found yet

*SCHEME 1. Choice set generation algorithm with stochasticity incorporated at attribute and traveler's preferences level.*

A specific challenge in the shortest path computation in a multi-modal network is to avoid illogical sequences of transport modes in the O-D trip. For example, given a certain multi-modal network and traveler group preferences the following paths may be generated if only looking for the minimum path cost:

- car-train-car;
- bike-train-bike.

However, observations show that car and bike are not used at the destination side unless the specific travel group has the realistic opportunity to use those transport modes. To prevent the generation of such routes in the choice set, access to car and bicycle networks should be allowed only in a limited area close to the home-end of the trip. In this way these transport modes (car and bike) can be used as access modes to train services or as main modes. In our approach a high penalty is given to the access links to car and bicycle networks that are in the area far away from the home address. Thus equation (2) will include the *Penalty*, which is zero for car and bicycle boarding links that are in the area close to the home-end, otherwise it is set to infinity.

$$c_a^s = \alpha_m^s * C_a + VOT^s * \beta_m^s * X_a + CK_m^s * D_a + Penalty \quad (3)$$

The question may be raised whether the proposed procedure indeed is able in principle to generate realistic route sets. As stated earlier, realistic implies that the subjective choice set should be included, while at the same time unrealistic and irrelevant routes are excluded. The remainder of the paper will focus on the first condition. With respect to the unrealistic routes it can be stated that although randomization might lead to extraordinary poor routes, that by selecting the best route from the randomized network precludes with a high probability the inclusion of unrealistic alternatives in the choice set. This assertion can be deduced from the findings in earlier studies (Ben-Akiva *et al.*, 1984, De La Barra *et al.*, 1993, and Fiorenzo-Catalano and Van Der Zijpp, 2001, Ramming, 2002, and Van Der Zijpp and Fiorenzo-Catalano, 2002).

#### 4. Case study

A case study has been carried out to demonstrate the application and performance of the proposed approach. The considered case is the corridor between the cities of Dordrecht and Rotterdam in the Netherlands, which are about 30 kilometers apart, with a total population of about one million. As the availability of private vehicles is clearly important in multi-modal route-choice it was decided to focus on home-based trips in which privately owned vehicles are available to travelers. Travelers in this corridor can use car and train as their main mode. In the case study three types of train services are available: local services, express services, and intercity services. Two stations in Dordrecht are considered in the corridor: Dordrecht Central, at which all services call, and one station served by local and express train services only (Dordrecht Zuid). Among all Rotterdam's railway stations four of them are considered in the corridor: Rotterdam Central (all services), Rotterdam Lombardijen (express and local services) and two stations served only by local trains: Rotterdam Zuid, and Rotterdam Blaak. All stations in the area are accessible by foot, bicycle, car, bus, tram and metro (the latter two in Rotterdam only). Both central stations have extensive facilities for bicycle storage and bicycle renting, but car-parking facilities at Rotterdam Central are limited. The resulting supernetwork consists of about 11,000 nodes and 34,000 links.

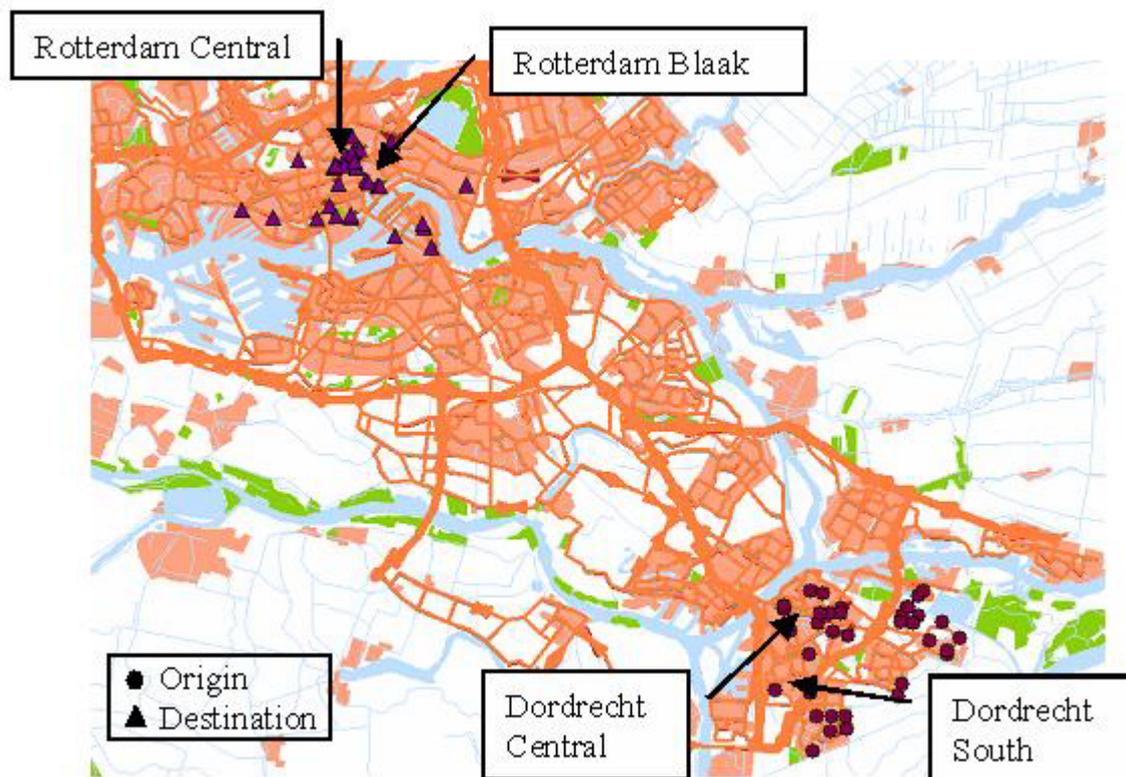


Figure 1. Overview of the corridor Dordrecht-Rotterdam and the selected trip origins and destinations

In order to analyze whether the generated route sets include the subjective choice sets, a comparison is made with observed route sets (Hoogendoorn-Lanser, 2005). For practical reasons, this analysis is limited to a set of 37 OD-pairs in the corridor Dordrecht (home-end) and Rotterdam (activity-end) during the morning peak hour (7.00 to 9.00). Figure 1 shows the locations of the origins and destinations of these trips. Please note that this analysis sets high standards for the comparison. The generation method is designed to generate route sets for a group of travelers, while the comparison is made with respect to individual choice sets.

Furthermore, an analysis is made of the impact of randomizing network attributes only, of randomizing travelers' preferences only, or of randomizing both components of the generalized costs. Therefore in the present paper the following three approaches are considered in turn:

1. Incorporation of stochasticity at network level by randomizing only the link attributes ( $X$ ) (random attribute approach);
2. Incorporation of stochasticity in the travelers' utility function by randomizing only the parameters ( $\beta$ ) (random preference approach);
3. Incorporation of stochasticity at both attribute level ( $X$ ) and traveler's preferences ( $\beta$ ) (combined randomization approach).

Compared to the approach shown in Scheme 1, in the random attribute approach steps (d) and (e) are skipped and the generalized link costs are computed with respect to the average values of the behavioral parameters. In the random preference approach only the behavioral parameters are randomized, steps (a) and (b) are skipped and the generalized link costs are computed with respect to the average values of link attributes.

#### **4.1 Dataset characteristics**

To evaluate the performance of our subjective choice set generation approaches, use has been made of observed trips collected from a large survey conducted among Dutch train travelers in 2001. This survey is part of a data collection program at the Delft University of Technology focusing on mode and route choice for inter-urban trips (Hoogendoorn-Lanser, 2005). In the telephonic interview a recently made trip, during which the traveler was screened, was discussed in great detail. Travelers reported their chosen alternative, that is, the sequence of transport modes and the transfer nodes, as well as alternatives they knew related to access modes, train service types, boarding or alighting stations and egress modes. These alternative routes are assumed to be representative for the subjective choice set. In the case of the selected 37 OD pairs the total number of observed routes is 67; the average size of the reported subjective choice sets is between two and three alternatives per OD-pair, with a minimum of one and a maximum of six reported alternatives per individual respondent.

#### **4.2 Application of the choice set generation algorithm**

To generate routes in the network, the choice set generation algorithm described in section 0 is adopted to compute the shortest paths with respect to the randomized generalized cost. The subjective choice sets are generated by repeatedly applying the choice set algorithm for a total of 1600 iterations per OD-pair: 16 traveler groups (4 trip purposes and 4 vehicle availability and vehicle preference categories) and 100 iterations for randomized attributes and parameters of the generalized cost function.

As noted in section 3, for practical reasons only the  $\beta$  parameters among the behavioral parameters ( $\alpha$ ,  $\beta$ ,  $VOT$ , and  $CK$ ) are randomized in the present approach. All parameter values ( $\alpha$ ,  $\beta$ ,  $VOT$ , and  $CK$ ), however, are varied based on traveler groups. Adopted values of the  $\beta$  parameters and their standard deviations are synthesized from route-choice models presented in literature, especially (Van Der Waard, 1988) (Table1).

**Table 1 Adopted Values for the  $\beta$  Parameters**

$\beta$ parameters	$E(\beta)$	$\sigma$	Minimum ( $E(\beta)-\sigma$ )	Maximum ( $E(\beta)+\sigma$ )	Variation (%)
PT in-vehicle time	1.0	0.7	0.3	1.7	71%
PT board/alight time	1.5	1.0	0.5	2.5	68%
PT waiting time	2.5	1.5	1.0	4.1	60%
In-vehicle time car	1.0	0.7	0.3	1.7	71%
Board/alight time car	1.0	0.7	0.3	1.7	71%
In-vehicle time bicycle	1.0	0.7	0.3	1.7	71%
Board/alight time bicycle	1.5	1.0	0.5	2.5	68%

The values used are in line with those found in more recent studies (see e.g. Nielsen, 1996 and Wardman, 2001). Please note, that all values used in this analysis are based on literature and experience, and are not estimated nor optimized for this application.

To generate sufficiently realistic route sets taking into account the variety of traveler's behavior the considered three randomization approaches are applied as follows:

1. Random attribute approach (RA): 100 randomizations of the link time attributes ( $X$ ) with steps (d) and (e) skipped and an average value for the behavioral parameters adopted in the cost function;
2. Random preference approach (RP): 100 randomizations of the weights ( $\beta$ ) for travelers' preferences for travel time components with steps (a) and (b) skipped and an average value for the link attributes adopted in the cost function;
3. Combined randomization approach (RC): 10 randomizations of the network time attributes ( $X$ ) each combined with 10 randomizations of the traveler's preferences ( $\beta$ ) for time implying that 10 is the number of iterations for the steps (a) and (d) respectively of the algorithm described in Scheme 1. The total number of iterations is thus again 100 equal to the number of iterations of the previous described approaches.

Table 2 summarizes some of the characteristic outcomes of the generation procedures. Routes are always unique, although may be overlapping to some extent. During the route generation procedure, all paths generated that are 100% overlapping with a previously found path are rejected (see step (h) of the algorithm described in Scheme 1). In order to compare the generated subjective choice sets with the sample of observed individual route sets additional constraints are used to account for the traveler's vehicle availability. Furthermore, since the observed route sets only contain train trips, routes by private modes only were also skipped from the analysis.

**Table 2. Choice Set Generation Results for Sample of 37 OD-Pairs**

	CS-RA	CS-RP	CS-RC	Theoretical extreme
Total number of distinct routes generated	286	283	701	1600 x 37
Average number of routes per choice set	8	8	19	1600
Coefficient Of Variation	52%	35%	32%	
Maximum Size	18	16	36	1600
Minimum Size	3	4	10	1

The choice set (CS) generated by the RA approach (CS-RA) contains 286 route alternatives with an average size of 8 alternatives per OD-pair, a minimum of 3 and a maximum of 18 route alternatives. The choice set (CS) generated by the RP approach (CS-RP) contains 283 route alternatives with an average size of 8 alternatives per OD-pair, a minimum of 4 and a maximum of 16 route alternatives. The choice set (CS) generated by the combined approach (CS-RC) contains 701 route alternatives with an average size of 19 alternatives per OD-pair, a minimum of 10 and a maximum of 36 route alternatives.

The sizes of these choice sets look plausible and do not conflict with empirical knowledge about objective choice sets (Bovy and Stern, 1990). Please note, that the average size (see Table 2) of the resulting choice sets is reasonable due to the fact that the generation method generates choice sets for a group of travelers (at aggregate level). Interestingly the first two generated choice sets (CS-RA and CS-RP) are quite similar in size. However, it might be expected that there are clear differences in route compositions. It can be hypothesized that the randomization of network attributes might lead to a smaller variety of routes than the randomization of behavioral parameters. It appears that the combination of randomized attributes and travelers' preferences generates much more different route alternatives with the same number of iterations. This might be explained by the fact that the routes generated by varying network attributes usually vary around the shortest path. If variations of travelers' preferences lead to more different routes, the combined approach would lead to additional variations on these different routes. Finally, the figures in Table 2 show that the generated number of routes is a lot less than the theoretical maximum. The randomization approaches generate a lot of identical or nearly equal alternatives being not included in the choice set.

## 5. Performance comparison of choice set generation

The key question in this section is how well the subjective choice sets generated by each approach match the observed route sets. In comparing two sets, e.g. A and B, we define the set coverage of A with relation to B as the percentage of alternatives in set A that are also elements of set B. We are looking for the coverage of the observed sets with relation to the generated sets. We distinguish three levels of comparison with increasing level of detail:

- Station level: home-end station and activity-end station combination;
- Leg level: home-end mode, train service types, activity-end mode;
- Trip level: unique combination of home-end mode, home-end station, train service type, activity-end station, and activity-end mode.

The subjective choice sets generated by each of the three approaches (CS-RA, CS-RP, and CS-RC) are compared with the observed alternatives: the chosen trip alternative (CA) and the set of trip alternatives (including the chosen alternative) reported to be known (KA) by the traveler (reported subjective choice set).

Table 3 shows the set coverage results for the choice sets (CS-RA, CS-RP, and CS-RC) generated by the three approaches and their comparison with the chosen alternative (CA) and the known alternatives (KA). As we can see from the table the best coverage at trip level is given by the combined approach RC with a percentage of 78%, which might seem obvious given the number of routes found using the RC approach. More interesting is that the RP approach in which the randomization is applied to the travelers' preferences produces a better coverage than the RA approach. Apparently, the variation in behavioral parameters is more important when considering route alternatives, than the variation in network attributes.

**Table 3. Set Coverage Results for Each of the Three Generated Choice Sets: CS-RA, CS-RP, CS-RC**

N=37 OD-trips	CS-RA		CS-RP		CS-RC	
	CA $\subseteq$ RA	KA $\subseteq$ RA	CA $\subseteq$ RP	KA $\subseteq$ RP	CA $\subseteq$ RC	KA $\subseteq$ RC
Home-end and activity-end railway stations	91.9 %	91.0 %	86.5 %	84.2 %	94.6 %	93.7 %
Home-end leg modes	54.1 %	54.9 %	97.3 %	96.2 %	97.3 %	96.9 %
Train leg	89.2 %	88.3 %	54.1 %	53.2 %	89.2 %	88.3 %
Activity-end leg modes	83.8 %	86.5 %	83.8 %	82.0 %	91.9 %	91.9 %
Complete trip	37.8 %	40.5 %	51.4 %	50.2 %	78.4 %	77.8 %

If we analyze the coverage results of the combined approach we can observe that: at the first level, home-end and activity-end station, the set coverage is very high for both the chosen alternatives and the known alternatives: about 94%. At the second level, individual legs, the set coverage is still high: more than 88% of the reported legs are part of the generated subjective choice set. At the trip level, the set coverage is still high: about 78% for both comparisons, even if it is not as high as the comparison with the trip components. Of course, the classification of high and low might seem arbitrary. The comparison of observed and generated route sets, however, is fairly new. In a recent study (Ramming, 2002) a comparison is made between observed route (chosen route only) and generated routes for a road network. A coverage was found of 72% for a combined labeling method, 60% for multiple-path algorithms, and 50% for a simulation method with optimized values for the standard deviation. Given these findings, the results of our method can be classified as promising.

There may be various reasons why the set coverage at trip level does not reach the same high percentage as at the trip components level. On the one hand it might be due to assumptions in modeling the transport system (for example, timed transfers in low frequency networks), while on the other hand it might be caused by atypical individual behavior. A third reason might be that a more detailed description of travel behavior with respect to trip composition is needed.

When we look at the coverage levels of the reported alternatives by the choice sets generated by the combined approach for the 37 OD-pairs, we can see that for 26 OD-pairs the set coverage is 100%. For 5 OD-pairs the reported route set is partly covered ranging between 80% and 20%. For 6 OD-pairs the set coverage is nil. If we look closer at these 6 cases it

appears that these all consist of only one reported alternative, being the chosen one. Furthermore, we see, in two cases, that all components are generated, but not the reported alternatives. In both cases the reported alternatives have a longer travel time: activity-end mode walking instead of tram or Express train instead of intercity train. However, the differences are relatively small. In other two cases, the travelers choose to use the local train even though the intercity train would bring them faster to their activity-end station. Apparently, there are some unaccounted benefits in using the local train service, such as maybe the seat availability.

This analysis of cases where the reported route set was not part of the generated choice set shows that the main reason can be found in the network description and the generation algorithm. Only in a few cases, atypical individual travel behavior explains why the reported alternatives could not be generated.

Given these findings the algorithm can be improved in many ways. First of all, a sensitivity analysis should be performed with respect to the variances at the link level as well as with respect to the weights ( $\beta$ ). The values used in this analysis were based on literature only. Given the difference in size between the generated choice sets and the reported choice sets, special attention should be given to reducing the size of the generated choice set. One option might be to have a more detailed utility-function that matches the traveler's behavior. An alternative approach might be to apply a kind of filter on the generated set in order to eliminate uninteresting routes for a specific traveler or group of travelers. However, please note that, the difference in size of the generated and the reported choice sets is mainly due to the fact that the choice set generation method aims at generating a route choice set for an OD-pair at the zone level, whereas the reported choice sets are observed for individual travelers. Finally, since the number of cases having a bad match between observed and generated is still quite high, more knowledge on actual travel behavior should be incorporated, such as preferences for specific modes or mode combinations.

## **6. Conclusions**

Multi-modal traveling involves complex alternatives consisting of multiple different legs forcing travelers to choose for transport services, modes, and boarding and alighting railway stations, etc. This paper describes an algorithm to generate route sets in multi-modal networks. To generate realistic route sets taking into account the variety of traveler's behavior three approaches were considered:

1. Incorporation of stochasticity for the link attributes used to model travelers' perception of the network;
2. Incorporation of stochasticity for the weights used to model travelers' preferences for trip attributes;
3. Incorporation of stochasticity at both attribute level and travelers' preferences.

The choice sets generated by each approach using a supernetwork of the corridor between the Dutch cities Rotterdam and Dordrecht were compared with a set of observed multi-modal route sets. All three approaches generate plausible routes and reasonable choice set sizes.

Comparing the generated choice sets with the observed routes leads to interesting conclusions:

- The best coverage of the observed routes is provided by the combined approach in which both link attributes and weights for travelers' preferences are randomized.
- The approach that incorporates stochasticity only for the weights that are used to model travelers' preferences provides better coverage than the approach that includes stochasticity only for network attributes.

A possible explanation for this finding is that when travelers think about alternative routes they start with the question: what could I do differently? This point of view is clearly related to travel behavior and not to the stochasticity in the network.

It proved to be possible to generate most of the trip components that make up an alternative. However, generating the complete alternatives proved to be more difficult, but the choice set generated by the combined approach still covered the set of the observed routes with a good percentage of 78%. The number of alternatives containing in these generated choice sets is much larger than those of the choices sets generated by the other two approaches, but it is still relatively large since they require about 700 alternatives generated to cover a set of observed routes of only 67 alternatives. There are however multiple options for performance improvement such as by the way how the network is modeled, e.g. timed transfers in low frequency networks, by introducing in the travelers' cost function some mode specific constants to deal with the different perception of specific transport modes and services, and the like.

Generating route choice alternatives prior to choice prediction offers a lot of benefits such as more flexibility in choice modeling and much less computing time in an iterative assignment application context. The analysis presented here has shown that a priori generation is a feasible approach for realistic multi-modal networks, while sufficiently realistic choice sets can be established.

## Acknowledgements

The research presented in this paper is part of the Hypernetworks project, which is partly funded by the joint research program of the Netherlands Research School for Transport, Infrastructure and Logistics (TRAIL) and the Netherlands Organization of Applied Scientific Research (TNO). This publication is also a product of the research program Seamless Multi-modal Mobility, carried out within TRAIL. This article is an adapted and improved version of a paper presented at 83<sup>rd</sup> Annual Meeting of the Transportation Research Board January 2004.

## References

Ben-Akiva, M., Bergman, M.J., Daly, A.J. and R. Ramaswamy (1984) Modeling inter-urban route choice behaviour, In: *Proceedings Ninth International Symposium on transportation and traffic theory*, (Volmuller, J., R. Hamerslag, eds.), VNU Science Press, Utrecht, Netherlands, pp. 299-330.

Ben-Akiva, M., and M. Bierlaire (1999) Discrete choice methods and their applications to short term decisions. *Handbook of transportation science*. Ed. R.W. Hall. Pp. 5-33. Kluwer Academic Publishers. Boston/Dordrecht/London.

- Bliemer, M.J.C., H.H. Versteegt, and P.H.L. Bovy (2003) INDY: A new analytical multiclass dynamic traffic assignment model, *Paper presented at the INFORMS Annual Meeting*, Atlanta, 2003
- Bliemer, M.J.C., E. Versteegt, and R. Castenmiller (2004) *INDY: A New Analytical Multiclass Dynamic Traffic Assignment Model*, Paper for TRISTAN V, Guadeloupe June 2004.
- Bovy, P.H.L., and E. Stern (1990) *Route choice: wayfinding in transport networks*. Kluwer Academic Publishers.
- Carlier, K., S. Fiorenzo-Catalano, K. Lindveld and P. Bovy (2003). A supernetwork approach towards multi-modal route choice modeling. Presented at the *82nd Annual Meeting of the Transportation Research Board*, Washington, DC., 2003, Paper 03-1951 on Preprint CDROM, 1-16.
- De La Barra, T. Perez B. and J. Anez (1993) Multidimensional path search and assignment. *Proceedings of the 21 PTRC Summer Meeting*, 1993.
- Fernandez, E., De Cea, J., Florian, M., and E. Cabrera. (1994) Network equilibrium models with combined modes, *Transportation Science*. Vol. 28, No. 3, pp. 182-192.
- Fiorenzo-Catalano, S., and N.J. Van Der Zijpp (2001). A Forecasting Model for Inland Navigation based on Route Enumeration. In: *European Transport Conference*, 10-12 September 2001, Cambridge, PTRC education and research services ltd. London, UK, 2001, pp.1-11.
- Hoogendoorn-Lanser, S. (2005) *Modeling travel behavior for multi-modal transport networks*, TRAIL Thesis Series T2005, TRAIL, The Netherlands (to be published)
- Hoogendoorn-Lanser, S. and R. Van Nes (2004), Multi-modal choice set composition: Analysis of reported and generated choice sets, Paper 04-3949 CD-ROM *83rd Annual Meeting of the Transportation Research Board*, 2004
- Nielsen, O.A. (1996) Do Stochastic traffic assignment models consider differences in road users' utility functions? *Proceeding of the 24th PTRC conference 1996 Seminar M*.
- Nielsen, O.A. (2002) Timetable-based transit assignment with error components in the utility-function, In: Lam W.H.K. (ed): *Proceedings Croucher workshop on Advanced modeling for transit operations and service planning*, The Hong Kong Polytechnic University, Hong Kong, 2002
- Nuzzolo, A., F. Russo, and U. Crisalli (2001). A doubly dynamic schedule-based assignment model for transit networks, *Transport Science* 35 (3). Pp. 268-285.
- Ramming, M.S (2002) *Network knowledge and route choice*. PhD. Thesis, MIT.
- Sheffi, Y. and W.B. Powell (1982) An algorithm for the Equilibrium Assignment Problem with Random Link Times, *Networks*, Vol. 12, No. 2, pp. 191-207
- Van Der Waard, J. (1998) *The relative importance of public transport trip-time attributes in route choice*, PTRC, 1988.

Van Der Zijpp, N.J. and M.S. Fiorenzo Catalano (2002). An Algorithm for Finding the K-Acyclic Shortest Paths with Logical Constraints and its Application to the Route Enumeration Problem, Presented at the *81th Annual Meeting of the Transportation Research Board*, Washington, DC., 2002, Paper 02-3981 on Preprint CDROM, 1-13

Van Nes, R. (2002) *Design of multi-modal transport networks, a hierarchical approach*, T2002/5, TRAIL Thesis Series, Delft University Press, The Netherlands 2002.

Wardman, M. (2001) *Public transport values of time*, Working paper 564, Institute of Transport Studies, University of Leeds, Leeds, 2001.