Some Properties and Implications of Stochastically Generated Route Choice Sets

Michiel C.J. Bliemer, Piet H.L. Bovy, Hao Li

Delft University of Technology, Faculty of Civil Engineering and Geosciences
Department of Transport & Planning, P.O. Box 5048, 2600 GA Delft, The Netherlands
Tel. +31.15.2784874, Fax. +31.15.2783179
Email: m.c.j.bliemer@tudelft.nl, p.h.l.bovy@tudelft.nl, h.li@tudelft.nl

1 Introduction

Choice sets of individual travelers play a paramount role in analyzing travel choice behavior. Choice sets are defined as the collection of travel options perceived available by individual travelers in satisfying their travel demand. From a variety of studies it is well known that the size and composition of choice sets do matter in cases of choice model estimation and demand prediction. Incorrect choice sets can lead to misspecification of choice models and to biases in predicted demand levels. While this has been demonstrated for relatively simple choice types such as mode choice, we may assume that it holds as well for the more complex case of route choice. The critical role of choice sets in choice modeling has given rise to profound research into choice set modeling in the transportation field, although largely confined to mode choice. We state that these insights gained on choice set modeling and choice set generation cannot simply be transferred to the route choice realm. For a variety of reasons, the specification of route sets for route choice modeling is different and more complex, reason why this topic deserves special attention from researchers and practitioners.

This paper will be devoted to a number topics related to the generation of route choice sets, some statistical properties of the size and composition of these sets, and the application of route choice sets in forecasts.

Different route choice set generation procedures exist. Some are constrained enumeration methods, based on certain decision rules, using so-called branch-and-bound algorithms to add routes to the choice set (e.g., Hoogendoorn-Lanser, 2005; Prato and Bekhor, 2006). Others are repeated (stochastic) shortest path methods which randomly add routes to the choice set (e.g., Fiorenzo-Catalano et al., 2004; Bovy and Fiorenzo-Catalano, 2006). In this paper the focus will be on the latter.

2 Stochastic route choice set generation

2.1 Theoretical description

Stochastic route choice set generation procedures typically assume probability distributions over route attributes and taste parameters. The utility function of each route is expressed as a linear combination of different route attributes weighted with taste parameters of the travelers:

$$\tilde{V}_p^{rs} = \sum_k \tilde{\beta}_k \tilde{X}_{pk}^{rs}, \tag{1}$$

where $\tilde{X}_{pk}^{rs}$ is the k-th attribute of route p from origin r to destination s, e.g., travel time, and $\tilde{\beta}_k$ is the associated weighting (taste) parameter. This utility function can be viewed as a route generation function, where each traveler aims to maximize his/her utility. The route attributes $\tilde{X}_{pk}^{rs}$ (like travel time) are usually computed from link attributes $\tilde{X}_{ak}$ such that $\tilde{X}_{pk}^{rs} = \sum_{a\in\Gamma_p^{rs}} \tilde{X}_{ak}$, where Γ_p^{rs} is the set of links constituting route p.

Let P^{rs} denote a population of unique alternative routes from r to s existing in a network, and let $\bar{P}^{rs} \subseteq P^{rs}$ denote a generated subset. The stochastic route choice set generation procedure determines routes to enter set $\bar{P}^{rs}$ by iteratively finding the cheapest (or shortest, or fastest) path, given instants of the random attributes and taste parameters. If for each iteration (with specific random draws) the cheapest route does not exist in route set $\bar{P}^{rs}$ yet, then it is added to this route choice set. This procedure is repeated N times. For more information, see Fiorenzo-Catalano (2006).

2.2 Example

Let us assume that the route generation function is simply defined by $\tilde{V}_p^{rs} = -\tilde{\tau}_p^{rs}$, where $\tilde{\tau}_p^{rs}$ denotes the (instantaneous) stochastic route travel time. Hence, the taste parameter is assumed to be fixed, only the attribute (travel time) is stochastic. This stochastic route travel time is assumed to be composed of stochastic link travel times $\tilde{\tau}_a$, such that $\tilde{\tau}_p^{rs} = \Sigma_{a\in\Gamma_p^{rs}} \tilde{\tau}_a$. Furthermore, the stochastic link travel time will be defined by

$$\tilde{\tau}_a = \tau_a^0 \left(1 + |\varepsilon_a|\right), \quad \text{with} \quad \varepsilon_a \sim N(0, \sigma^2), \tag{2}$$

where τ_a^0 is the uncongested link travel time. Note that $\tilde{\tau}_a \geq \tau_a^0$, such that $\tilde{\tau}_a$ could be viewed as some kind of congested travel time. The standard deviation parameter σ can be used to specify the range of travel times a link can attain. If $\sigma = 0$, then only the fasted route will be found, while if $\sigma > 0$ more routes will be generated. Alternative specifications for the stochastic link travel times and the route generation function can be used, this specification is merely used for illustration purposes.

Consider two simple example networks, an "urban" grid network depicted in Figure 1, and a "motorway" network in Figure 2 where typically major cities have direct motorway connections. In each network, we are interested to generate a route choice set for trips from origin *r* to destination *s*. Assume that all bi-directional links in the network have an equal uncongested travel time of $\tau_a^0 = 1$. For such small networks we could easily list all possible routes from *r* to *s*, as listed in Figures 1 and 2. For the "urban" network, routes 1–6 have an uncongested route travel time of 4, routes 7–10 have an uncongested route travel time of 6, and routes 11–12 have an uncongested route travel time of 8, respectively. In the "motorway" network, these uncongested route travel times are 2, 3, 4, and 5, respectively for route 1, routes 2–4, routes 5–8, and routes 9–10.

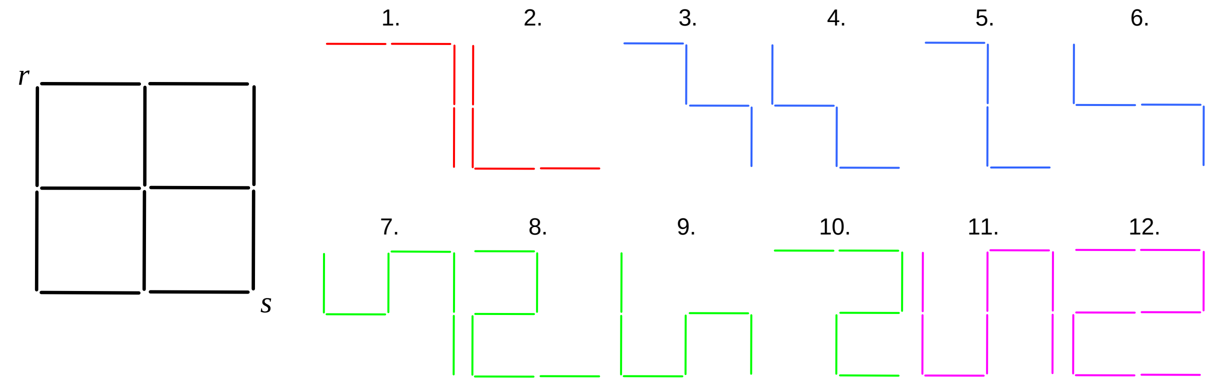


Figure 1: Example "urban" network (left) and all twelve possible routes from *r* to *s* (right).

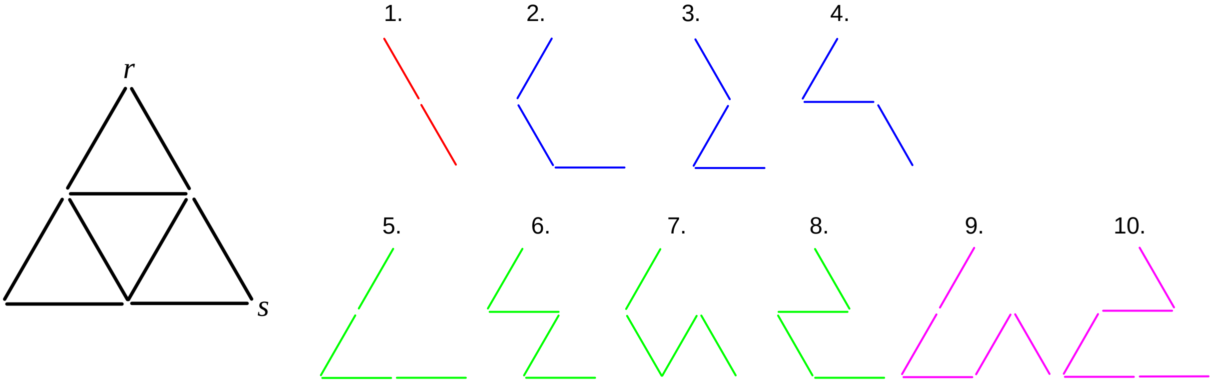


Figure 2: Example "motorway" network (left) and all ten possible routes from *r* to *s* (right).

As mentioned before, we are typically not interested in enumerating all routes, but only the routes that are most likely to be considered in the route choice set. For the “urban network” we would be mostly interested in generating routes 1–6, and for the “motorway” network mainly route 1. All other routes are less likely to be chosen as they have an uncongested route travel time of at least 50% extra. Given a certain σ, the selection probabilities of finding each route using the stochastic choice set generation method are plotted in Figure 3. They have been computed by 10,000 Monte Carlo simulations. Note that these selection probabilities have no connection with route choice proportions, they merely indicate the probability of each route to be included in the route choice set. If one would like to have a larger route choice set, it is best to have all selection probabilities as close together as possible by changing the route generation function. Some interesting observations can be made from Figure 3.

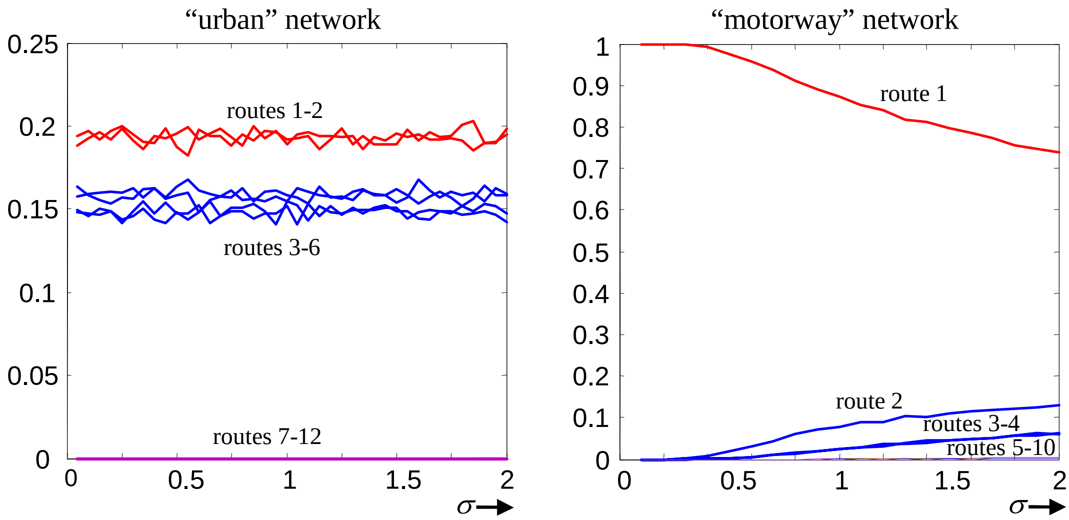


Figure 3: Selection probabilities of finding each route in the “urban” and “motorway” network

First consider the “urban” network. Even though routes 1–6 all have an equal uncongested travel time, routes 1–2 have a higher probability of being generated, which may seem counterintuitive at first sight. However, it can be explained as follows. Looking at routes 1–6, routes 1 and 2 use some links that are not used by any of the other routes, namely the 2nd and 3rd links. If the link travel times on these links are small in a certain random draw, then only route 1 or 2 profits from this and are likely to be the fastest route. In case any of the other links have a small link travel time, there are always at least two routes that profit, hence they are competing over being the fastest route in which one of those routes will not be generated for that random draw. Hence, routes 1 and 2 have a slightly higher probability of being generated. Also note that routes 7–12 will never be generated, independent of the standard deviation σ, which can be explained by the fact that all these routes share at least three of the same links with routes 1–6. For example, for a certain random draw of link travel times, it is very unlikely that route 7 will be faster than route 1.

Now consider the “motorway” network. In contrast to the “urban” network, the probabilities of finding each route depends on σ. The larger this standard deviation, the higher the probability of a longer route to be faster, which will typically yield a larger route choice set. Route 1 is clearly the fastest route in the network and will therefore have a much higher probability of being generated. With a low σ, only route 1 will be generated, while with higher σ also route 2 and perhaps even routes 3 and 4 will be generated. Routes 5–10 are difficult to generate.

Obviously, the choice for the route generation function and the distribution of the route or link attributes strongly determines the generated routes. Although not shown here, if instead of $\tilde{\tau}_a = \tau_a^0 (1 + |\varepsilon_a|)$ one would have used $\tilde{\tau}_a = \tau_a^0 (1 + \varepsilon_a)$, the probabilities of choosing a long route would increase, therefore yielding larger choice sets.

3 Statistical properties of the choice set

3.1 Theoretical description

In the previous section, the probability of a route being chosen is concerned, which can be viewed as the well-known coupon collectors problem (Von Schelling, 1954). Here we can define the probability of a route being chosen as the selection probability η_p^{rs} of a route p from origin r to destination s. Due to stochastic properties of the route generation function in both variations in the taste parameter and the route attributes, the resulting route choice set in terms of size and composition, and the number of random draws required to generate a predetermined route choice set are stochastic variables. It becomes of interest to capture the stochastic properties of the choice set and try to establish theoretical supports that can govern the experimental set up of the generation process.

Omitting the indices (r,s), let N denote the number of route selections (which is equal to the number of iterations of the route set generation algorithm), let Q denote the size of set P^{rs}, and let K denote the size of subset $\bar{P}^{rs}$. Furthermore, define R_K as the number of route selections from P^{rs} required to establish for the first time a subset $\bar{P}^{rs}$ of size K. Define K_N as the size of subset $\bar{P}^{rs}$ after N route selections from $\bar{P}^{rs}$. Finally, define N_p as the number of route selections from P^{rs} required to have for the first time a particular route p into subset $\bar{P}^{rs}$.

We derived four interesting properties which can be categorized into three groups: sample size-related, choice set size-related and composition-related properties. Sample size-related properties are specifically for given set size with unknown draws, while choice set size-related properties are for given number of selections with unknown resulting set size. Composition-related properties can provide information of the composition of the generated choice set and especially for a specific route of interest. The properties are mathematically formulated:

<u>Sample size-related properties:</u>

1. Assuming that the routes between a certain OD pair have approximately equal selection probabilities, The cumulative inclusion probability of having K different routes in $\bar{P}^{rs}$ after N route selections is given by

$$\Pr(R_K \le N) = \sum_{m=K}^{N} \Pr(R_K = m), \text{ with } \Pr(R_K = m) = \sum_{j=0}^{m-K-1} \Pr(R_K = m-1-j) \cdot \sum_{r=1}^{K-1} \left(\frac{r}{Q}\right)^{j+1} \cdot \frac{1}{(m-K)} \quad (3)$$

With this formula, the required number of route selections to achieve a choice set with given size k with a certain confidence level can be derived.

2. The number of route selections N required to achieve a high confidence (at least 90%) of generating an exhaustive choice set (such that $K = Q$) is about 5 times of the population size, i.e. $R_Q^{0.90} \approx 5Q$. This indication could be used as an easy rule of thumb.

<u>Choice set size -related properties:</u>

3. After N route selections, the average number of (distinct) routes generated in set $\bar{P}^{rs}$ is

$$E(K_N) = Q - \sum_{p \in P^{rs}} (1-\eta_p^{rs})^N \quad (4)$$

<u>Composition-related properties:</u>

4. The cumulative probability of having a particular alternative route $p \in P^{rs}$ in choice set $\bar{P}^{rs}$ after N route selections is

$$\Pr(N_p \le N) = 1 - \Pr(N_p > N) = 1 - (1-\eta_p^{rs})^N \quad (5)$$

This property is specially for the interest in a specific route, which due to different purposes is desired to be generated in the choice set.

Properties 1 and 2 are specially derived for equal selection probability case. However they are accurate enough for the unequal probability case where all the alternative routes are more or less equally attractive (see Li et al., 2006).

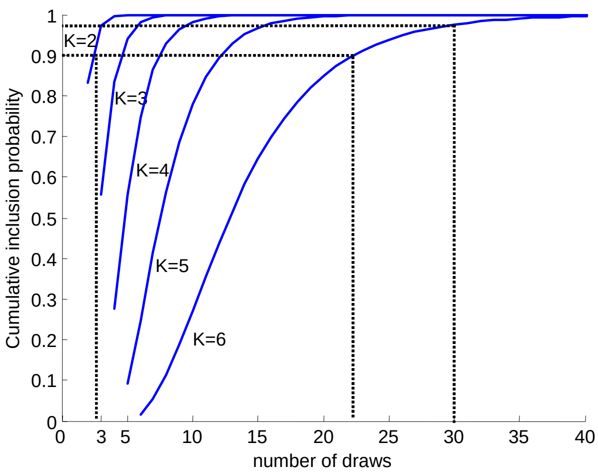
3.2 Example

The theoretical derivations listed above will be applied to the "urban" network. Some interesting questions are the following:

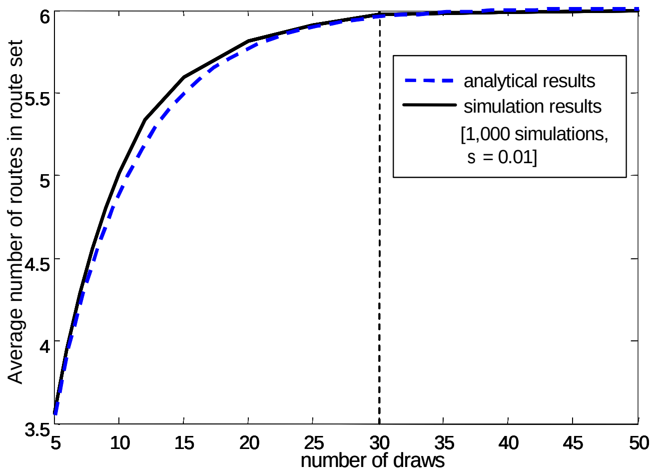
(a) How many route selections are needed in order to generate the (six) most attractive routes with a 90% confidence level?

(b) What is the probability that a certain route will be included in the route choice set when a pre-specified number of route selections is performed?

The six most attractive routes in the "urban" network have approximately equal selection probabilities. Figure 4(a) depicts the cumulative inclusion probabilities of Equation (3) and Figure 4(b) depicts the average size of the generated route set from both analytical analysis and simulation-based analysis, which gave consistent results (see Equation (4)). Answering question (a), assuming a population size $Q = 6$, about $N = 22$ route selections are needed to guarantee with a 90% confidence level that $K = 6$ routes are included (see property 1 and Figure 4(a)). If only $K = 2$ routes are needed, only 3 draws are sufficient to have a very high confidence level. If $N = 30 = 5 \cdot 6$ (conform property 2) route selections are performed, then with more than 90% confidence an exhaustive choice set will be generated. Figure 4(b) indicates that 30 route selections are necessary to obtain on average six routes in the choice set, which is consistent with the analytical results from property 3. Answering question (b), suppose there is an interest in route 1. If $N = 10$, then according to property 4, with $1-(1-0.19)^{10} \approx 88\%$ confidence level route 1 will be included in the route choice set, while if $N = 5$, this is only 65%.



(a)



(b)

Figure 4: Cumulative inclusion probability and average choice set size for "urban" network

4 Impact of route choice set on route choice probabilities

Ideally, route choice models should be robust to the route choice set under consideration. In other words, adding route alternatives to the choice set that are unlikely to be chosen should not change the route choice probabilities much, at least not the relative choice proportions. However, in many route choice models this is not the case, as will be illustrated below.

Consider again the "urban" example network in Figure 1, and suppose we apply a route choice model based on two different route choice sets: (a) a route choice set containing only the six most attractive routes. $\bar{P}_1^{rs} = \{1,\ldots,6\}$, and (b) a complete route choice set consisting of all twelve possible routes, $\bar{P}_2^{rs} = P^{rs} = \{1,\ldots,12\}$.

Using the uncongested travel times, we compute the route choice proportions ψ_p^{rs} for each route *p* from origin *r* to destination *s* using the multinomial logit (MNL) model and the path-size logit (PSL) model. The PSL model, introduced by Ben-Akiva and Bierlaire (1999) and extended by Ramming (2002) and Hoogendoorn-Lanser (2005), is defined by

$$\psi_p^{rs} = \frac{\exp\left(\mu(-\tau_p^{rs} + \theta \ln \xi_p^{rs})\right)}{\sum_{p' \in \bar{P}^{rs}} \exp\left(\mu(-\tau_{p'}^{rs} + \theta \ln \xi_{p'}^{rs})\right)}, \quad \text{with} \quad \xi_p^{rs} = \sum_{a \in \Gamma_p^{rs}} \left(\frac{\tau_a^0}{\tau_p^{rs}}\right) \frac{1}{\sum_{p' \in \bar{P}^{rs}} \delta_{ap'}^{rs} \left(\frac{\tau^{rs,*}}{\tau_{p'}^{rs}}\right)^{\gamma}}, \tag{6}$$

where δ_{ap}^{rs} equals one if link *a* is on path *p* from *r* to *s* and zero otherwise, $\tau^{rs,*}$ is the fastest path from *r* to *s*, and μ, θ, and γ are parameters. The MNL model is a special case in which the path-size factor ξ_p^{rs} is equal to one for all paths *p* (or $\theta = 0$). The MNL model cannot handle route overlap, as it assumed that all route alternatives are independent, while the path-size factor in the PSL model explicitly accounts for this. The route choice proportions for routes 1–6 are listed in Table 1 using $\mu = 1$ and $\theta = 1$.

Table 1: Route choice proportions for different route choice sets and models

		route											
route set	**model**	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{P}_1^{rs} = \{1,\ldots,6\}$	MNL	0.17	0.17	0.17	0.17	0.17	0.17	-	-	-	-	-	-
	PSL, all γ	0.22	0.22	0.14	0.14	0.14	0.14	-	-	-	-	-	-
$\bar{P}_2^{rs} = \{1,\ldots,12\}$	MNL	0.15	0.15	0.15	0.15	0.15	0.15	0.02	0.02	0.02	0.02	0.00	0.00
	PSL $(\gamma = 0)$	0.12	0.12	0.16	0.16	0.16	0.16	0.02	0.02	0.02	0.02	0.00	0.00
	PSL $(\gamma = 1)$	0.13	0.13	0.15	0.15	0.15	0.15	0.03	0.03	0.03	0.03	0.00	0.00
	PSL $(\gamma = 2)$	0.15	0.15	0.14	0.14	0.14	0.14	0.03	0.03	0.03	0.03	0.01	0.01
	PSL $(\gamma = 10)$	0.07	0.07	0.05	0.05	0.05	0.05	0.15	0.15	0.15	0.15	0.03	0.03

As expected, the choice proportions for the MNL model for routes 1–6 are identical, since these routes all have equal uncongested travel time and route overlap is not taken into account. The lower route choice proportions in the MNL model using $\bar{P}_2^{rs}$ compared to $\bar{P}_1^{rs}$ is merely due to the fact that $\bar{P}_2^{rs}$ contains more route alternatives (routes 7–12) that have small nonzero choice proportions. The route choice proportions with $\bar{P}_1^{rs}$ using the PSL show the difference that route overlap makes, making routes 1 and 2 more attractive than routes 3–6 (which is consistent with our explanation in Section 2). However, if we include all routes as in $\bar{P}_2^{rs}$, then the PSL model predicts *lower* choice proportions for routes 1 and 2 (independent of the values chosen for μ, and θ) compared with routes 3–6. Adding unattractive routes such as routes 11 and 12 increases the route overlap on the attractive routes 1 and 2, such that the route choice proportions of the latter go down. In general, adding (long and unattractive) routes that overlap with attractive routes decrease the choice proportions of the attractive routes due to the path-size factor in the PSL model. Hence, the PSL model is not robust. A way to overcome this problem may be to weight the path-sizes corresponding to the attractiveness (length) of the route, which is similar to what has been proposed in the *exponential* PSL model (Ramming, 2002) in which $\gamma > 1$. Indeed, when $\gamma > 1.9$, routes 1 and 2 become more attractive again. When γ is chosen too high (e.g., $\gamma = 10$), then the path-size factors become too dominant such that routes 7–10 will become most attractive. Hence, the exponential PSL model can be very sensitive to the choice of γ.

5 Conclusions

In this paper stochastic route set generation models are considered and illustrated using some simple but illustrative examples. Stochastic procedures are typically able to generate the most attribute routes, while unattractive routes are automatically excluded from the choice set.

Route choice models are in general sensitive to the route choice set (although ideally they should not). Unattractive routes can create biases in the route choice proportions. Either the route choice model should be adapted in order to avoid these biases, or the route choice set generation model should only generate more or less attractive routes. Stochastic route set generation models as discussed in this paper are able to distinguish between these attractive and unattractive routes in a very simple and fast fashion. Settings for the route choice set generation model can be derived from the statistical properties and discussed, giving an indication of how many iterations may be necessary in order to generate all more or less attractive route alternatives. Furthermore, these generation models are able to implicitly take the network structure into account, such that depending on the network (being e.g. urban or motorway) different sizes and compositions of the route set result.

References

Ben-Akiva, M., and M. Bierlaire (1999) Discrete Choice Methods and their Applications to Short Term Travel Decisions. In R. Hall (ed.) *Handbook of transportation science*, Kluwer.

Bovy, P.H.L., and S. Fiorenzo-Catalano (2006) Stochastic route choice set generation: Behavioural and probabilistic foundations. *Paper presented a the 11th International Conference on Travel Behaviour Research*, Kyoto, Japan.

Fiorenzo-Catalano, S., R. van Nes, and P.H.L. Bovy (2004) Choice set generation for multi-modal travel analysis. *Proceedings of the TRB Annual Meeting*, Washington DC, USA.

Fiorenzo-Catalano (2006) Choice set generation in multi-modal transport networks. PhD Thesis, Delft University of Technology, The Netherlands (forthcoming).

Hoogendoorn-Lanser, S. (2005) *Modelling travel behaviour in multi-modal networks*. PhD Thesis, Delft University of Technology, Delft, The Netherlands.

Li, H., P.H.L Bovy, G. Hooghiemstra (2006) Probabilistic properties of stochastic choice set generation. Working report, Delft University of Technology, Transportation and Planning Section.

Prato, C.G., and S. Bekhor (2006) Applying brach & bound technique to route choice set generation. *Proceedings of the TRB Annual Meeting*, Washington DC, USA.

Ramming, M.S. (2002) *Network knowledge and route choice*. PhD Thesis, MIT, Cambridge, USA.

Von Schelling, H. (1954) Coupon collecting for unequal probabilities. *American Math. Monthly*, Vol. 61, pp. 304-311.