Simulation-based generation of route choice sets in large public transport networks

Thomas Kjær Rasmussen, Otto Anker Nielsen & Carlo Giacomo Prato *

Department of Transport,

Technical University of Denmark, Kgs. Lyngby, DK-Denmark

* Email: {tkra, oan, cgp}@transport.dtu.dk

1 Introduction

In recent years, research within route choice modeling has pointed towards an approach where the choice set is generated prior to model estimation and flow prediction. This has led to increasing attention towards the importance of size and composition of choice sets for traffic route choice (see [1,2] for overviews). The generation of such route choice sets for public transport has not been investigated much in the literature, which has mainly devoted its attention to route choice of car users.

This study presents a method of generating route choice sets for public transport networks, and analyses its sensitivity to the stochasticity of its parameters. Specifically, this paper describes the generation of choice sets for users of the Greater Copenhagen public transport system by applying a doubly stochastic path generation algorithm, and then evaluates them in comparison with actual public transport route choices collected in the Danish Travel Survey.

2 Methods

This study analyzed 4783 observations of actual route choices in the Greater Copenhagen public transport network. The data was collected within the Danish Travel Survey through a detailed questionnaire gathering all relevant trip information about feeder mode, routes, lines, departure and travel times and purposes [3]. Given the observed public transport routes, corresponding choice sets were generated using a doubly stochastic simulation method that allows for detailed specification of the utility function of alternative i [4,5,6]:

$$\begin{aligned} U_i = \beta_{walktime} \cdot TT_{walktime,i} + \beta_{waittime} \cdot TT_{waittime,i} + \beta_{changepen,i} \cdot N_{change,i} \\ + \beta_{conntime} \cdot TT_{conntime,i} + \beta_{waitzone} \cdot TT_{waitzone,i} + \beta_{IVT,train} \\ \cdot TT_{IVT,train,i} + \beta_{IVT,ICtrain} \cdot TT_{IVT,ICtrain,i} + \beta_{IVT,S-train} \cdot TT_{IVT,S-train,i} \\ + \beta_{IVT,bus} \cdot TT_{IVT,bus,i} + \beta_{IVT,metro} \cdot TT_{IVT,metro,i} + \varepsilon_i \end{aligned}$$

where $TT_{walktime}$ and $TT_{waitime}$ is walking and waiting time when changing, N_{change} is the number of changes, $TT_{conntime}$ is time spent walking between the origin/destination and the first/last public transport stop, $TT_{waitzone}$ is time spent waiting at the origin, and $TT_{IVT,train}$, $TT_{IVT,ICtrain}$, $TT_{IVT,S\text{-}train}$ and $TT_{IVT,bus}$ are in-vehicle times spent in regional trains, intercity trains, suburban trains and buses respectively. The corresponding β's are parameters and ε_i is an error term.

Path sets were generated for three different configurations of the utility function:

- *ErrTermOnly*: β's are not randomly distributed across the population, and ε_i is Gamma distributed.
- *ErrCompAll*: β's are Log-Normal distributed, and ε_i is not considered in the utility function.
- *ErrCompErrTerm*: β's are Log-Normal distributed, and ε_i is Gamma distributed.

The specification of these three configurations are based on work done in Rasmussen([7]). Each configuration was investigated with nine levels of stochasticity (i.e., 27 specifications). The stochasticity was introduced by using parameters that scale the variance of the distribution: the variance of the β's and ε_i is $\beta_{parameter}*\mu_{parameter}$ and $\beta_{error\ term}*V_{deterministic,i}$, respectively[1]. The nine levels of the scale parameters, also based on findings in Rasmussen([7]), are shown in Table 1, and *ErrCompErrTerm_7* thus refers to a specification where β's and ε_i are distributed at a scale of 1.5. Tests on the levels denoted by _1, _2 and _3 showed that performance increases with increasing size of the scale parameters [7]. Consequently, to find the level from which the performance ceases to improve but worsens with increasing stochasticity, this study additionally tested cases with higher scale parameters. β's were drawn at OD-level in each iteration, whereas ε_i's were drawn at link level for each OD in each iteration.

Level of stochasticity	_1	_2	_3	_4	_5	_6	_7	_8	_9
$\beta_{parameter}$	0.05	0.10	0.15	0.2	0.5	1	1.5	2	5
$\beta_{error\ term}$	0.025	0.05	0.10	0.2	0.5	1	1.5	2	5

Table 1- Size of parameter that scale the variance of the distribution

[1] For the gamma distribution, the traditional formulation with shape k and scale θ can be computed by: $\Gamma(k,\theta) = \Gamma\left(\left(\frac{\mu}{\sigma}\right)^2, \frac{\sigma^2}{\mu}\right)$. For the Log-Normal distribution, the traditional formulation with log-scale A and shape B^2 can be computed by: $LogN(A,B^2) = LogN\left(ln(\mu) - 0.5 * ln\left(1+\frac{\sigma^2}{\mu^2}\right), ln\left(1+\frac{\sigma^2}{\mu^2}\right)\right)$

For each specification, the path set was generated by performing 200 iterations and then identifying the unique routes. For these, the overlap in length with the recorded route choice was computed, and the coverage, equal to the share of observations for which a certain overlap exists between the observed and generated paths, could then be calculated [8].

3 Results

Ideally, the stochastic generation technique should produce a variety of relevant unique routes within a reasonable number of iterations. But with a high level of stochasticity, the simulation seemed to continue to generate new unique paths (Figure 1). More thorough analysis showed, however, that after a number of iterations, redundant and non-intuitive routes were generated. This indicates that it is undesirable to iterate until the number of unique routes stabilizes.

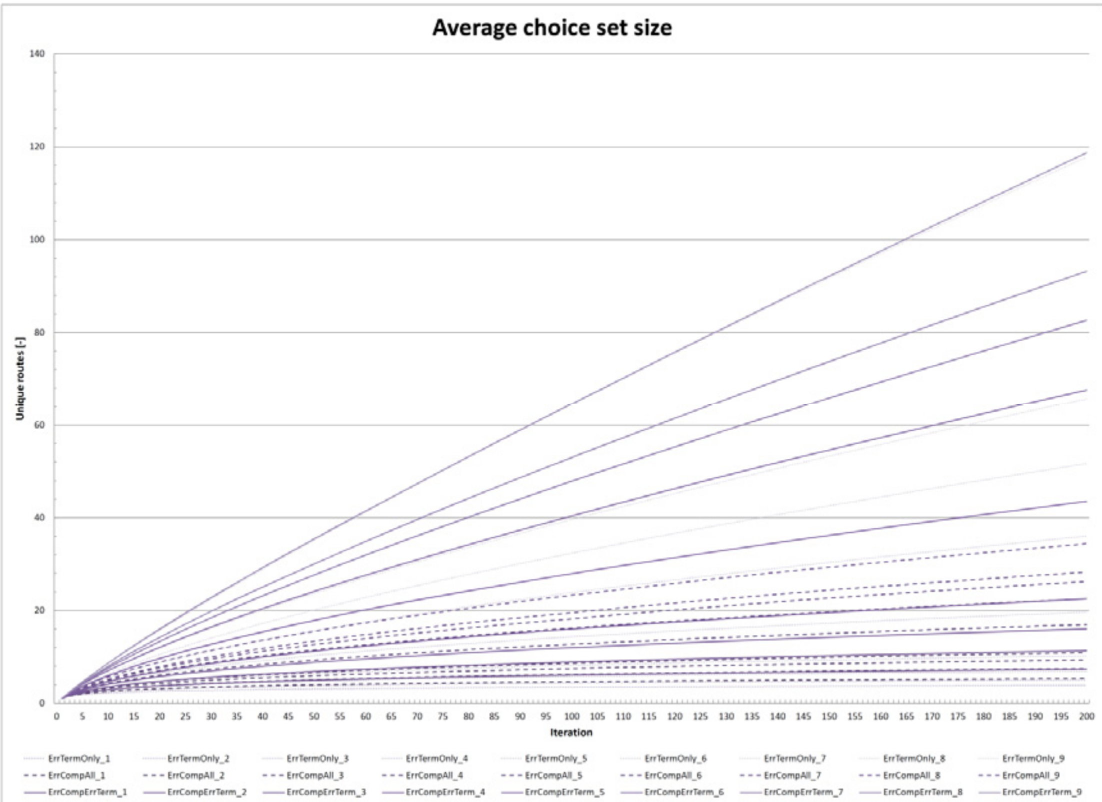

Figure 1 - Choice set size

Figure 2-3 show, for two trips, the observed route and the choice set generated by configuration *ErrCompErrTerm_6*. Redundant and counterintuitive routes were generated in both cases (e.g. loops with the metro can be seen in Figure 3).

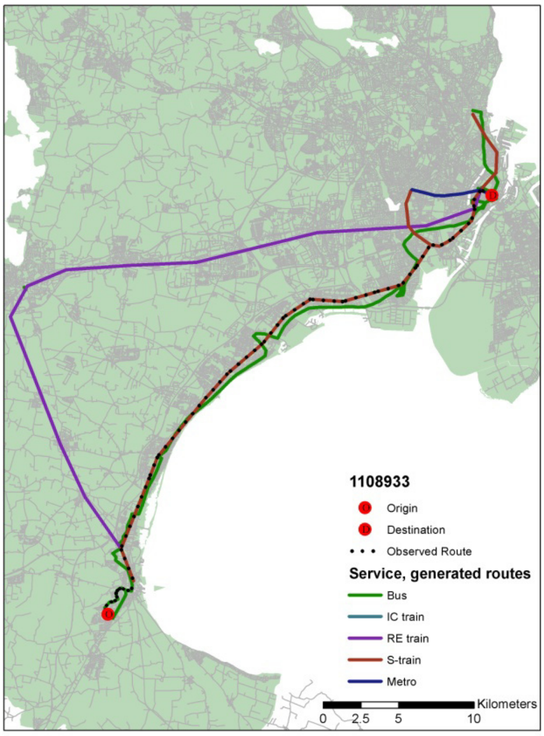

Figure 2 - Choice set and observed route , I

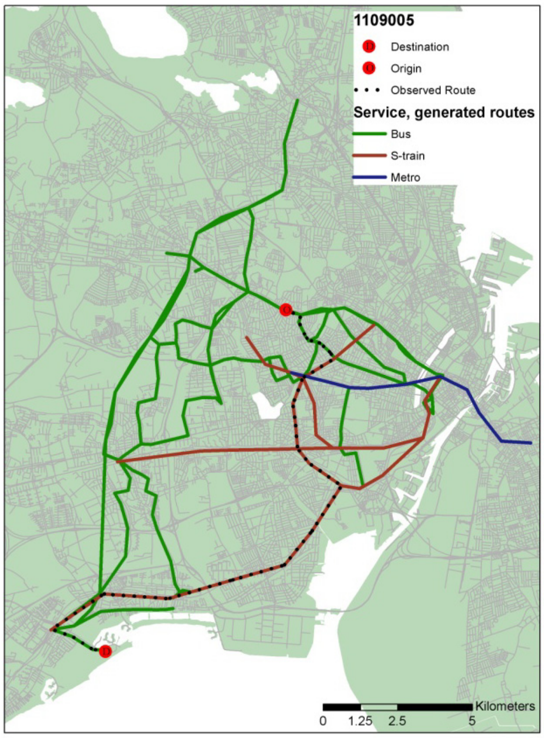

Figure 3 - Choice set and observed route , II

As an alternative to convergence in the number of unique routes, coverage improvement could be used as an indicator of performance. An overlap threshold of 80% produced the results illustrated in Figure 4 for configuration *ErrCompErrTerm*.

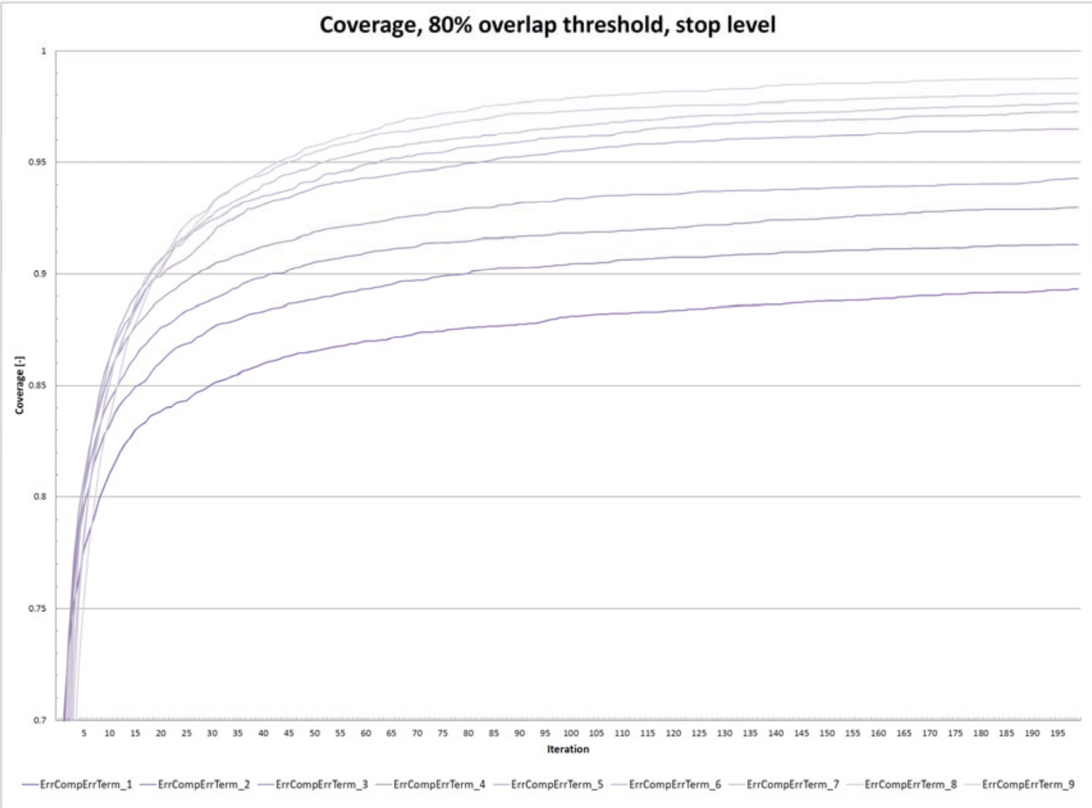

Figure 4 - Coverage, configuration *ErrCompErrTerm* with overlap threshold of 80%

The results suggest that the coverage grows when the parameters are randomly distributed across the population. The best specification was the combined model where parameters and the error term were distributed.

The results for all specifications were considerably improved by including a distributed change penalty, which indicates that most travelers actually do associate a penalty to changes.

It was observed that a certain amount of stochasticity is needed in order to generate a representative choice set with high coverage without needing a lot of iterations. But adding very high stochasticity is however not desired either – too many redundant and counterintuitive routes are generated.

4 Conclusions

This study is among the first to investigate actual route choices of public transport users and assess choice set quality against these. The method applied was found suitable for choice set generation in a large-scale network, and the study clearly highlights the importance of the algorithm and utility-specification chosen for the choice set generation. It was found that the level of the introduced stochasticity should be done with parsimony, as adding stochasticity translates into the generation of redundant and counterintuitive routes after a certain level. Adding variability across people improved the results considerably, and as models are going in the direction of being more disaggregate in nature, it is important for them to account for as much individual variability as possible.

References

[1] Bovy, P. H. L., “On modelling route choice sets in transportation networks: a synthesis”, Transport Reviews 29(1), 43-68, 2009

[2] Prato, C. G., “Route choice modeling: past, present and future research directions”, Journal of Choice Modelling 2(1), 65-100, 2009

[3] Larsen, M. K., “Development and assessment of a data collection method for route choice in public transport”, conference proceedings for Aalborg Trafikdage, Aalborg, 27-28 August, 2010

[4] Nielsen, O. A., “A stochastic transit assignment model considering differences in passengers utility functions, Transportation Research Part B 24(5), 377-402, 2000

[5] Nielsen, O. A., “A large scale stochastic multi-class schedule-based transit model with random coefficients”, In: Wilson, N., and Nuzzolo, A. (Eds.) “Schedule-Based Dynamic Transit Modelling – Theory and Applications, Kluwer Academic, pp. 51-77, 2004

[6] Nielsen, O. A., Frederiksen, R. D., “Optimization of timetable-based, stochastic transit assignment models based on MSA”, Annals of Operations Research 144(1), 263-285, 2006

[7] Rasmussen, T. K., “Public Transport route choice in the Copenhagen Region – generation and assessment of route choice sets” (in Danish), Master Thesis at Department of Transport, Technical University of Denmark, 2010

[8] Ramming, M. S., “Network Knowledge and Route Choice”, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2002