DTU Library

Judgments in the selection of path generation techniques: a meta-analytic approach

Prato, Carlo Giacomo

Published in:
The 91st Annual Meeting of the Transportation Research Board, Washington,DC 2012

Publication date:
2012

Judgments in the Selection of Path Generation Techniques: A Meta-Analytic Approach

Carlo Giacomo Prato *
Technical University of Denmark
Department of Transport
Bygningstorvet 116 Vest
2800 Kgs. Lyngby, Denmark
Telephone: +45.45256595
Fax: +45.45936533
E-mail: cgp@transport.dtu.dk

* Corresponding author

Revised paper submitted for presentation at the 91th Annual Meeting of the Transportation Research Board, January 22-26, 2012, Washington D.C.

Submission date: November 12, 2011

Word count: 5,330 (text) + 10 × 250 (figures and tables) = 7,830

ABSTRACT

In the context of modeling route choice behavior, modelers generate objective choice sets by selecting a path generation technique and its parameters according to personal judgments. The current paper proposes an experimental setting and a methodological approach to provide indications about objective judgments for effective route choice set generation.

Initially, path generation techniques are implemented within a synthetic network to generate possible subjective choice sets considered by travelers. Next, "true model estimates" and "postulated predicted routes" are assumed from the simulation of a route choice model. Then, objective choice sets are applied for model estimation and results are compared to the "true model estimates". Last, predictions from the simulation of models estimated with objective choice sets are compared to the "postulated predicted routes".

Meta-analysis allows synthesizing the effect of judgments on the implementation of path generation techniques, since a large number of models generate a large amount of results that are otherwise difficult to summarize and to process. Meta-analysis estimates suggest that modelers should apply stochastic approaches with the possibility of correcting for unequal sampling probability while maintaining a fairly reasonable level of variance. Estimation of models would greatly improve and the issue of the coverage of observed behavior would not be raised because the correction would cover for adding alternatives not generated.

1. INTRODUCTION

In the context of modeling route choice behavior in transport networks, travelers are presumed to choose the best alternative from a set of routes connecting origin and destination of their trip. Although most equilibrium assignment techniques implicitly assume that all existing paths between origin and destination are available to travelers, conceptual and empirical reasons suggest that explicit path generation prior to discrete choice model estimation or path-based traffic assignment is preferable.

Conceptually, choice set formation and choice from alternatives are distinct mental processes calling for separate modeling: choice set formation is trial-and-error determined (*1*), preference-driven (*2*) and constraint-related (*3*), while choice from alternatives is usually represented as a compensatory decision (see *4*, *5*). Empirically, various case-studies show advantages of explicit choice set formation: higher flow prediction accuracy is illustrated for path-based solutions to the Stochastic User Equilibrium (SUE) problem (*6*), unrealistic and inefficient paths are found within implicit choice sets for link-based assignment (*7*), and theoretical and computational advantages are shown when choice set generation is performed prior to traffic assignment (*8*).

Several solutions have been proposed to the explicit path generation problem. Deterministic solutions include variations of shortest path algorithms (e.g., *9*, *10*, *11*), minimization of generalized cost functions (*12*), application of heuristic rules combined with shortest path searches (e.g., *13*, *14*), and implementation of a branch-and-bound algorithm (*15*). Stochastic solutions include single stochastic simulation (e.g., *16*, *17*), doubly stochastic simulation (*18*, *19*) and a random walk algorithm (*20*). Advantages and disadvantages related to the implementation of existing path generation techniques are extensively discussed by Bovy (*4*) and Prato (*5*).

Even though several solutions have been proposed to the explicit path generation problem, guidelines for the implementation of path generation techniques have never been provided. As modelers cannot observe the subjective choice sets that contain the routes considered by travelers, they generate objective choice sets by selecting a path generation technique and its parameters according to personal judgments.

Although the impact of choice sets on choice probabilities and model performances has received increasing attention recently, the effect of path generation techniques on model estimates and flow predictions has not been documented. Model performances from the estimation of route choice models with two different choice sets generated with (i) branch-and-bound algorithm and (ii) merged deterministic techniques have been compared in terms of likelihood values (*21*, *22*). Unfortunately, the comparison fails to evaluate which technique better represents the observed behavior because of the absence of information with respect to actual values of model estimates. Choice probabilities from several route choice models for a synthetic network with three choice sets containing respectively 6, 10 and 12 alternatives have been compared to choice probabilities from a postulated probit model (*23*). Unfortunately, the analysis fails to assess the effects of the implementation of path generation techniques because of the peculiar context of a universal realm of only 12 alternatives. Convergence properties have been examined for SUE problem solutions with different objective function values and choice set sizes (*24*). Unfortunately, the comparison fails to investigate path generation techniques other than the k-shortest path algorithm. In a nutshell, existing studies about choice set effects on route choice models focus on the robustness of models and methods, rather than on the effects of path generation techniques and on the provision of general indications about judgments for generating choice sets.

The current study proposes an alternative approach. Initially, subjective choice sets possibly considered by travelers are generated by implementing path generation techniques within a synthetic network. Next, "true model estimates" and "postulated predicted routes" are assumed from the simulation of a route choice model. Then, objective choice sets are applied for model estimation and results are compared to the "true model estimates". Last, predictions from the simulation of models estimated with objective choice sets are compared to the "postulated predicted routes". The advantage of this approach is threefold: (i) assuming subjective choice sets according to behavioral assumptions behind various path generation techniques allows covering a large variety of possible behavior in the absence of any indication about actual subjective choice sets considered by travelers; (ii) estimating the same model specification within the same synthetic network allows isolating choice set effects from model and network effects; (iii) analyzing the possible combinations of subjective and objective choice sets allows comparing the relative ability of path generation techniques in accurately generating "postulated chosen routes" and reproducing "true model estimates".

The appraisal of the coverage of the postulated behavior with the objective choice sets and the assessment of the effects of path generation techniques on estimation and prediction accuracy are performed with a meta-analytic approach. Even though meta-analysis is generally used to review findings across different empirical studies, this paper proposes the application of meta-analysis to synthesize the effect of judgments when a large number of models generate a large amount of results that are otherwise difficult to summarize and to process. Judgments concern the path generation technique to be implemented and its parameters to be defined, and the meta-analysis examines a large number of combinations of subjective and objective choice sets to provide modelers with general requirements for obtaining better coverage of postulated behavior and higher accuracy in model estimates and flow predictions.

The remainder of the paper is structured as follows. Section 2 presents the rationale behind the consideration of path generation techniques for constructing subjective and objective choice sets. Section 3 describes the synthetic data and the methods for evaluating model estimates and flow predictions. Section 4 synthesizes estimation and prediction results. Section 5 summarizes the findings from the analysis.

2. GENERATING SUBJECTIVE AND OBJECTIVE CHOICE SETS

Judgments about path generation techniques and their parameters are examined by considering a variety of techniques that are included in the analysis according to the following four considerations.

The first consideration concerns the distinction between deterministic and stochastic techniques. Even though intuitively superior, considering only stochastic approaches would bias the current analysis by providing answers before questions about path generation effectiveness are even formulated.

The second consideration involves the selection of deterministic approaches. Even though their evolution suggests the superiority of more recent developments with respect to shortest path algorithms, considering only one deterministic technique would again bias the current analysis. The first deterministic technique considered is the most straightforward approach to the choice set generation problem consisting in the computation of *K*-shortest paths, as this technique is relevant because still largely applied in practice. The second deterministic technique considered is the iteration of the shortest path search after heuristic rules penalize links on the last shortest path computed in the iterative process, with link penalty (*14*) preferred to link elimination (*13*) to avoid network disconnection problems. The third deterministic technique considered is the enumeration of the paths connecting origin and destination of a trip under behavioral and logical constraints (*15*).

The third consideration concerns the selection of stochastic approaches. The current analysis examines the effects of three techniques most recently developed and used in route choice modeling. The first stochastic technique considered is the most straightforward stochastic simulation approach to the path generation problem consisting in the iteration of the shortest path search after randomization of link impedances (e.g., *17*, *19*). The second stochastic technique considered is the natural evolution of the previous approach considering a generation function with an error term for the traffic network variations and an error term for traveler taste heterogeneity (e.g., *18*, *19*). The third stochastic technique considered is a random walk algorithm that is biased towards the search for the shortest path (*20*).

The fourth consideration playing a role in the selection of path generation techniques involves the parameters for their implementation being relevant for model estimation and model implementation. For the k-shortest path, five values of *K* are selected to cover from a fairly small to a very large window of admissible path costs. For the link penalty, five combinations with increasing penalizing factor are chosen to cover from a small to a large variation of the alternatives in the generated choice sets. For the branch-and-bound, five combinations of thresholds of the branching rule (*15*) are defined in order to assess the effect of increasing choice set size and route heterogeneity in the generation process. For the stochastic simulation, a gamma distribution (see *18*) is preferred over normal distribution (see *16*, *25*) and truncated normal distribution (see *15*, *26*). Five combinations of shape and scale parameters are selected in order to have a mean equal to the link impedance and a range of increasing variances. For the doubly stochastic simulation, a gamma distribution is chosen for the error component representing impedance variation at the link level, and a log-normal distribution is selected for the error component representing heterogeneity in travelers' utility functions. Five variations are considered in order to cover a range of growing variances in both the perception of the traffic network and the preferences of the travelers. For the random walk algorithm, five values of the parameters of the Kumaraswami distribution of the weights (*20*) are accounted for in order to encompass a range of variance with respect to the shortest path in the generation process.

Summarizing, five variations are considered for each of six path generation techniques in order for the analysis to account for judgments in terms of selection of the type of technique, application of the specific technique, definition of the level of variance of the parameters, and generation of small or large choice sets.

3. EXPERIMENTAL SETTING

3.1. Synthetic data

The experimental setting applies six path generation techniques to the synthetic network represented in figure 1 that consists of 38 nodes and 64 links, with link length proportional to the length of the figure and some links having speed bumps. The network is originally a part of a real network of the city of Borlänge (Sweden) and has been presented by Frejinger et al. (*20*). The universal realm consists of 170 alternative routes between origin *O* and destination *D*, among which 29 have equal minimum length.

Subjective choice sets are unknown and hence their definition is hypothesized according to the behavioral assumptions behind the six path generation techniques applied (see *5*). Table 1 reports parameters of the five variations for each path generation technique. Parameters are defined from very small to very large variance, where very small (large) variance suggests that the resulting choice set is fairly small (large) since increasing the number of iterations or the variance of distribution parameters does not produce (produces) additional routes.

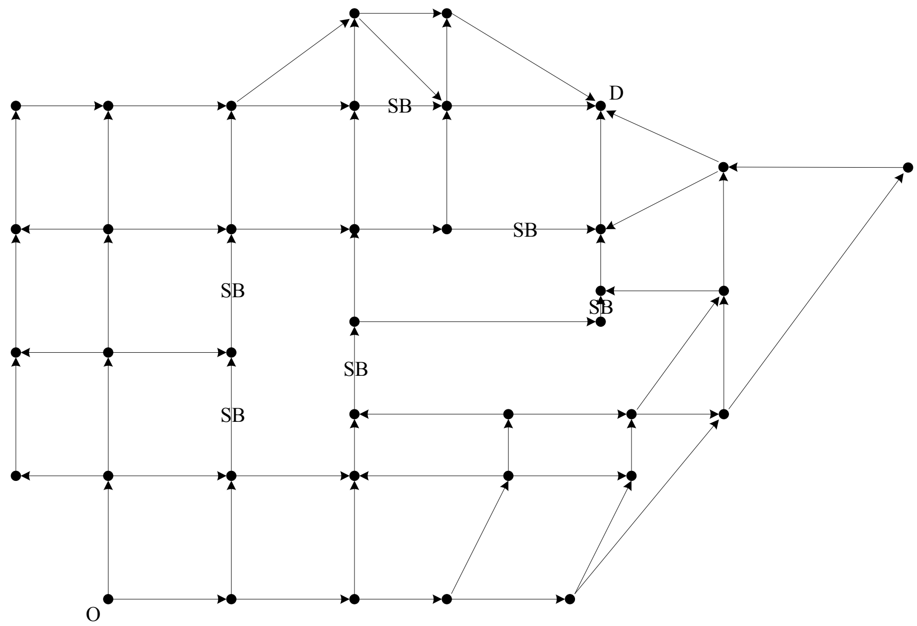

FIGURE 1 Network for experimental design (source: *20*).

Datasets of 4,000 observations for estimation purposes and 1,000 observations for prediction purposes are generated from the five variations of the six path generation techniques. This procedure results into 30 datasets of 4,000 subjective choice sets for model estimation and 30 datasets of 1,000 subjective choice sets for model prediction.

For each dataset of subjective choice sets, a PSC-Logit model (*28*) is postulated. The advantage in using the PSC-Logit is twofold: (i) the model accounts for similarities across alternatives while maintaining a simple Logit structure; (ii) MNL modifications are robust with respect to variations in the number of alternatives and in the composition of the choice sets (e.g., *22*, *23*).

The following utility function is specified for each alternative *j* and observation *n*:

$$U_{jn} = \beta_{length} Length_j + \beta_{bumps} SpeedBumps_j + \beta_{turns} Turns_j + \beta_{PSC} PSC_j + \varepsilon_{jn} \tag{1}$$

where $Length_j$ is the length, $SpeedBumps_j$ is the number of speed bumps, $Turns_j$ is the number of turns and PSC_j is the Path Size Correction of alternative *j*. The "true model estimates" are assumed equal to -1 for β_{length}, -0.10 for β_{bumps}, - 0.30 for β_{turns}, 1 for β_{PSC}, and error terms ε_{jn} are independently and identically distributed extreme value with scale 1 and location 0. The Path Size Correction of alternative *j* is defined as (*28*):

$$PSC_j = -\sum_{a \in \Gamma_j} \left(\frac{L_a}{L_j} \ln \sum_{l \in C} \delta_{al} \right) \tag{2}$$

where L_j is the length of route *j*, L_a is the length of link *a*, Γ_j is the set of links belonging to route *j*, *and* δ_{al} is the link-path incidence dummy (equal to one if links *a* belongs to route *l* and zero otherwise).

For each dataset of subjective choice sets, 4,000 "postulated chosen routes" for estimation purposes and 1,000 "postulated predicted routes" for prediction purposes are simulated by selecting the alternative with the highest utility within the choice set of each observation *n*.

TABLE 1 Parameters of Path Generation Techniques

Path generation technique	parameter	very small variance	small variance	average variance	large variance	very large variance
k-shortest path	define length limit in order to define k-shortest paths					
	length limit	35	36	37	40	51
	k	33	51	88	104	170
link penalty	iterate shortest path searches and penalize shortest path links					
	penalty factor	2%	3%	5%	10%	20%
	iterations	50	50	50	100	100
branch-and-bound	connect origin and destination of the trips with five behavioral and logical thresholds					
	directional	100%	100%	110%	110%	110%
	temporal	17	25	33	44	62
	detour	100%	100%	110%	110%	120%
	similarity	85%	80%	80%	75%	70%
	movement	4	4	4	5	5
stochastic simulation	iterate shortest path searches after extracting link length from gamma distribution					
	Gamma mean	length	length	length	length	length
	Gamma st.dev	0.25 length	0.50 length	length	2 length	3 length
doubly stochastic simulation	iterate shortest path searches after extracting link length from gamma distribution and travelers' preferences from log-normal distribution					
	Gamma mean	length	length	length	length	length
	Gamma st.dev	0.50 length	0.50 length	length	length	2 length
	LogN mean	-1	-1	-1	-1	-1
	LogN st.dev	0.25	0.25	0.5	1	1
random walk	calculate route probabilities from link probabilities based on link weights that are Kumaraswami distributed with two parameters b_1 and b_2					
	b_1	10	7	5	3	1
	b_2	1	1	1	1	1

3.2. Evaluation of model estimates and flow predictions

Objective choice sets are generated for model estimation and model prediction purposes from the 30 applied variations of path generation techniques.

For each dataset of objective choice sets, 30 models are estimated while considering as chosen alternatives the "postulated chosen routes" from the 30 datasets of subjective choice sets. Observations where the objective choice sets do not contain the "postulated chosen routes" are not considered for model estimation, and the coverage of the "postulated chosen routes" is evaluated for each of the 900 objective-subjective combinations:

$$Cov_{obj-sub} = \frac{\sum_n I\left(R_{sub,n} \in C_{obj,n}\right)}{N} \times 100 \qquad (3)$$

where $Cov_{obj\text{-}sub}$ is the coverage of the dataset *obj* of objective choice sets with respect to the "postulated chosen routes" from the dataset *sub* of subjective choice sets, $I(\cdot)$ is an indicator function equal to 1 when the "postulated chosen route" $R_{sub,n}$ belongs to the objective choice set $C_{obj,n}$ of observation *n*, and *N* is the total number of observations.

Models are estimated by defining the deterministic part of the utility function:

$$V_{jn} = \mu\left(\beta_{length} Length_j + \beta_{bumps} SpeedBumps_j + \beta_{turns} Turns_j + \beta_{PSC} PSC_j\right) \tag{4}$$

where μ is the scale parameter. It should be noted that β_{length} is fixed to -1 and μ, β_{bumps}, β_{turns} and β_{PSC} are estimated in order to have the same scale for all models and to compute the t-test with respect to the corresponding "true model estimates". When the random walk algorithm is used to generate datasets of objective choice sets, the deterministic part of the utility function is:

$$V_{jn} = \mu\left(\beta_{length} Length_j + \beta_{bumps} SpeedBumps_j + \beta_{turns} Turns_j + \beta_{PSC} PSC_j\right) + \ln\left(k_{jn}/q_j\right) \tag{5}$$

where k_{jn} is the number of time route j is sampled for observation n, q_j is the probability of sampling route j, and the logarithmic term corrects for the unequal sampling probabilities of the routes.

Accuracy of the model estimates with respect to the "true model estimates" of the postulated model is calculated for each of the 3,600 estimated parameters:

$$Acc_{par,obj-sub} = Prob\left(\frac{est_{par,obj-sub} - expest_{par}}{stderrest_{par,obj-sub}} < t\right) \times 100 \tag{6}$$

where $Acc_{par,obj\text{-}sub}$ is the accuracy of the estimate of the parameter for the model with dataset *obj* of objective choice sets and "postulated chosen routes" from dataset *sub* of subjective choice sets, $est_{par,obj\text{-}sub}$ is the estimate, $stderrest_{par,obj\text{-}sub}$ is its standard error, $expest_{par}$ is the expected "true model estimate", and t is the critical value of the Student distribution with n degrees of freedom.

For each dataset of subjective choice sets, Monte-Carlo simulation is applied by applying the estimates of the 30 models using the same dataset. The obtained "simulated predicted routes" are compared to the "postulated predicted routes" after translating both into network flows by counting the number of travelers on each link. Prediction accuracy is evaluated with the calculation of the following error measures for each combination of estimated model and dataset of subjective choice sets:

$$RMSE_{sim-pos} = \sqrt{\frac{\sum_{a=1}^{A}\left(N_{sim,a} - N_{pos,a}\right)^2}{A}} \tag{7}$$

$$MAPE_{sim-pos} = \frac{1}{A}\sum_{a=1}^{A}\left|\frac{\left(N_{pos,a} - N_{sim,a}\right)}{N_{pos,a}}\right| \tag{8}$$

where $RMSE_{sim\text{-}pos}$ is the root mean square error and $MAPE_{sim\text{-}pos}$ is the mean absolute percentage error between simulated and predicted routes, A is the number of links in the network, $N_{sim,a}$ is the flow on link a as calculated by translating the "simulated predicted routes", and $N_{pos,a}$ is the flow on link a as calculated by translating the "postulated predicted routes".

Given the 900 coverage values, the 3,600 accuracy values from model estimation, and the 900 mean absolute percentage errors, meta-analysis considers coverage, estimation accuracy and prediction error as dependent variables and characteristics of the choice sets as independent variables. Characteristics of the objective choice sets for model estimation include the technique applied and the degree of variance, while characteristic of the subjective choice sets for obtaining "postulated chosen routes" comprise choice set size (i.e., small for less than 30 alternatives, medium for 30 to 50 alternatives, large for more than 50 alternatives), degree of heterogeneity across routes (i.e., homogeneous for average path size less than 0.10), and consistency with the objective choice sets (i.e., both generated with the same path generation technique and same parameters).

4. RESULTS

4.1. Subjective and objective choice sets

Table 2 summarizes the characteristics of the choice sets from the implementation of the 30 variations of path generation techniques. Expected trends are that (i) the increase in the variance of the parameters for each path generation technique produces larger choice sets, (ii) deterministic techniques generate the same alternative routes for the same origin-destination pair, an unreasonable characteristic when considering that most likely different travelers have different subjective choice sets, and (iii) stochastic techniques produce different alternative routes for the same origin-destination pair, a desirable feature that seems behaviorally more plausible.

TABLE 2 Summary of Characteristics of Generated Choice Sets

Path generation technique	measure	very small variance	small variance	average variance	large variance	very large variance
k-shortest path	min	33	51	88	104	170
	max	33	51	88	104	170
	mean	33	51	88	104	170
	st.dev.	-	-	-	-	-
link penalty	min	22	29	43	53	49
	max	22	29	43	53	49
	mean	22	29	43	53	49
	st.dev.	-	-	-	-	-
branch-and-bound	min	17	25	33	44	62
	max	17	25	33	44	62
	mean	17	25	33	44	62
	st.dev.	-	-	-	-	-
stochastic simulation	min	24	28	33	44	49
	max	38	44	56	72	76
	mean	30.9	35.2	43.8	57.5	62.3
	st.dev.	4.3	5.8	8.3	11.0	11.4
doubly stochastic simulation	min	24	28	34	37	47
	max	40	46	59	64	73
	mean	31.7	35.7	45.6	49.9	58.6
	st.dev.	4.6	6.1	8.7	9.6	11.1
random walk	min	25	27	33	36	44
	max	42	50	55	66	71
	mean	33.2	38.5	44.0	51.8	58.7
	st.dev.	6.0	7.8	9.1	10.8	11.6

Table 3 summarizes the coverage of the datasets of objective choice sets with respect to the "postulated chosen routes". As 900 comparisons are computed, the table summarizes the relationship of the 30 variations of the path generation techniques with respect to "postulated chosen routes" by combining results for the five variations of each technique. Expectedly, regardless of the postulated behavior, (i) enlarging the number of alternatives considered in the k-shortest path and relaxing the thresholds in the branch-and-bound algorithm increases the coverage, (ii) the branch-and-bound outperforms competing deterministic techniques, and (iii) doubly stochastic simulation outperforms competing stochastic techniques. Unexpectedly, increasing the variance and enlarging objective choice sets does not boost the coverage. Possibly, the snapshot loses part of the information when considering jointly all five variations of each technique, and larger variance of the parameters of path generation techniques generates irrelevant routes that are not created under different parameters.

TABLE 3 Coverage of Objective Choice Sets with Respect to "Postulated Chosen Routes"

Path generation technique	variation	k-path	link penalty	branch and bound	stochastic simulation	doubly stochastic simulation	random walk
k-shortest path	very small var	54.7	41.0	50.0	42.2	42.2	41.3
	small var	82.0	71.7	76.9	70.8	70.8	69.7
	average var	94.7	87.3	87.2	87.3	87.3	86.8
	large var	97.6	91.3	92.2	93.0	95.6	91.2
	very large var	100.0	100.0	100.0	100.0	100.0	100.0
link penalty	very small var	73.5	69.2	71.1	66.8	66.7	65.2
	small var	61.1	75.8	66.7	64.7	64.4	62.6
	average var	79.4	84.8	83.6	75.0	74.5	73.8
	large var	72.4	85.1	78.0	75.3	75.4	73.6
	very large var	61.4	77.1	70.1	62.8	62.8	61.5
branch-and-bound	very small var	59.6	61.7	73.7	56.0	55.4	54.3
	small var	76.4	77.6	91.0	73.6	73.5	72.1
	average var	81.2	82.3	93.2	77.4	77.4	75.8
	large var	87.2	92.0	97.3	87.3	86.9	85.4
	very large var	99.5	99.9	100.0	99.7	99.7	99.1
stochastic simulation	very small var	95.5	96.4	96.5	96.4	95.8	94.9
	small var	94.9	95.8	96.1	96.3	95.4	94.4
	average var	93.1	93.8	94.4	94.8	93.5	92.5
	large var	85.6	87.6	87.6	88.9	86.5	85.5
	very large var	80.0	82.7	82.5	85.0	81.2	80.3
doubly stochastic simulation	very small var	95.7	96.4	96.7	95.9	96.8	95.0
	small var	94.8	96.0	96.3	95.4	96.5	94.4
	average var	92.3	93.3	93.7	92.8	94.1	91.9
	large var	90.5	92.0	92.1	91.1	92.8	90.4
	very large var	84.4	87.1	86.7	85.7	88.2	84.9
random walk	very small var	84.8	87.2	87.2	86.5	86.0	88.5
	small var	83.0	85.5	85.6	85.8	84.4	87.6
	average var	79.5	82.7	83.1	82.0	81.7	85.3
	large var	73.9	77.5	77.3	76.8	76.0	80.7
	very large var	59.2	65.7	65.4	64.5	63.5	71.0

The difficulty in interpreting coverage results from table 3 motivates the meta-analysis for the 900 comparisons. Table 4 presents estimates of the regression model that suggest how coverage of postulated behavior increases with the implementation of stochastic techniques, average to large variance of their parameters, and obviously application of the same technique for generating objective and subjective choice sets. Among deterministic techniques, branch-and-bound and link penalty contribute increasing coverage with respect to the k-shortest paths. Among stochastic techniques, the doubly stochastic simulation contributes augmenting coverage with respect to the stochastic simulation and even more the random walk. If the finding for deterministic techniques is expected, as the increase in coverage agrees with the growth in realism of the behavioral assumptions, the finding for stochastic techniques is less expected, as the more recently developed random walk does not outperform stochastic simulation. Coverage of "postulated chosen routes" benefits also from the analysis of large and homogeneous subjective choice sets, suggesting that path generation techniques perform better when the subjective choice sets are numerous and the alternatives are similar.

Meta-analysis estimates suggest that results from the snapshot of the coverage in table 3 might indeed be unexpected not only because of actual characteristics of the choice sets, but also because of results aggregated in the attempt to summarize findings from a large amount of models. The proposed approach allows not only considering every single combination of datasets of objective and subjective choice sets, but also suggesting general judgments in the implementation of path generation techniques regardless of the postulated behavior.

TABLE 4 Meta-Analysis Estimates of the Coverage

Parameter	est.	t-stat
characteristic related to the technique used to generate choice sets		
deterministic technique [a]	-	-
stochastic technique	26.718	23.76
k-shortest path [a]	-	-
link penalty	22.826	20.18
branch-and-bound	27.185	24.04
stochastic simulation	4.418	5.61
doubly stochastic simulation	8.060	8.87
random walk [a]	-	-
small variance	-10.896	-20.04
average variance [a]	-	-
large variance	4.064	7.63
characteristic related to the technique used to postulate choices		
low heterogeneity [a]	-	-
medium/high heterogeneity	-26.576	-30.99
small choice set size	-8.345	-10.64
medium choice set size [a]	-	-
large choice set size	7.655	15.77
consistent with generation	16.478	17.20
constant	39.122	62.36
N		3600
R^2		0.751

Notes: [a] reference category.

4.2. Accuracy of parameter estimates

Table 5 illustrates model estimates for the "postulated chosen routes" from the dataset of subjective choice sets corresponding to complete path enumeration (i.e., k-shortest path with 170 routes). This snapshot allows initial considerations about model estimates. Firstly, all the variations of the random walk algorithm allow obtaining unbiased model estimates, most likely because of the correction for unequal sampling probabilities of routes. None of the competing path generation techniques allows obtaining unbiased model estimates consistently, and only large to very large variance in their parameters allows reproducing the "true model estimates", suggesting that a larger generated choice set helps increasing estimation accuracy. Secondly, link penalty, branch-and-bound and both stochastic approaches fail almost in every circumstance to reproduce the "true model estimates", suggesting that obtaining higher coverage is not synonym of having higher accuracy in model estimation. Thirdly, k-shortest path shows some promise, but most likely because the "postulated chosen routes" are generated with a k-shortest path algorithm rather than for actual higher accuracy. Last, the parameter β_{bumps} seems to be recovered almost consistently, while the scale parameter μ appears to be recovered only sporadically.

TABLE 5 Snapshot of Model Estimates for Objective Choice Sets Generated from Different Path Generation Techniques

k-shortest path		very small var		small var		average var		large var		very large var	
parameter	value	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat
μ	1.00	0.556	13.1	0.627	11.7	0.780	7.4	0.969	1.2	0.972	1.1
β_{bumps}	-0.10	-0.132	-1.2	-0.127	-1.2	-0.101	0.0	-0.106	-0.3	-0.107	-0.3
β_{turns}	-0.30	-0.137	5.6	-0.092	13.0	-0.288	0.9	-0.283	1.2	-0.283	1.2
β_{PSC}	1.00	1.250	4.1	1.100	2.5	0.967	-0.8	0.943	-1.4	0.942	-1.4
LL(0)			-9430.7		-12894.6		-15648.7		-18420.7		-20543.2
LL(β)			-7284.0		-10369.2		-12311.0		-13471.8		-13472.3
rho-bar²			0.227		0.196		0.213		0.268		0.344

link penalty		very small var		small var		average var		large var		very large var	
parameter	value	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat
μ	1.00	0.626	13.9	0.714	10.9	0.953	1.7	0.782	8.6	0.979	0.8
β_{bumps}	-0.10	-0.182	-3.9	-0.129	-1.3	-0.501	-14.9	-0.181	-3.5	-0.415	-13.0
β_{turns}	-0.30	-0.134	10.8	-0.261	2.6	-0.092	15.3	-0.249	3.9	-0.062	17.1
β_{PSC}	1.00	0.757	-6.1	0.475	-13.0	0.389	-14.5	0.509	-11.9	0.252	-18.8
LL(0)			-13875.3		-14134.3		-15063.8		-15897.2		-15594.6
LL(β)			-12688.1		-12244.1		-12021.2		-12142.4		-11591.2
rho-bar²			0.085		0.133		0.202		0.236		0.256

branch-and-bound		very small var		small var		average var		large var		very large var	
parameter	value	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat
μ	1.00	1.300	-9.6	0.872	4.9	0.853	5.7	0.834	6.3	0.796	7.7
β_{bumps}	-0.10	-0.620	-14.1	-0.469	-18.1	-0.400	-14.2	-0.261	-7.7	-0.105	-0.2
β_{turns}	-0.30	-0.530	-10.7	-0.458	-7.6	-0.460	-7.4	-0.060	15.9	-0.281	1.4
β_{PSC}	1.00	-1.050	-43.3	0.256	-17.2	0.399	-14.1	0.465	-12.9	0.924	-1.9
LL(0)			-11441.2		-12923.5		-13367.1		-15151.7		-16509.2
LL(β)			-8997.5		-10413.5		-10575.6		-11303.3		-11863.4
rho-bar²			0.213		0.194		0.209		0.254		0.281

stochastic sim		very small var		small var		average var		large var		very large var	
parameter	value	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat
μ	1.00	0.392	22.5	0.578	15.6	0.729	10.1	0.841	5.9	0.873	4.7
β_{bumps}	-0.10	-0.097	0.1	-0.099	0.0	-0.103	-0.1	-0.110	-0.5	-0.118	-0.8
β_{turns}	-0.30	-0.264	2.6	-0.263	2.6	-0.261	2.8	-0.257	3.1	-0.259	2.9

β_{PSC}	1.00	0.851	-3.7	0.846	-3.8	0.826	-4.3	0.751	-6.2	0.715	-7.1
LL(0)			-13718.3		-14239.5		-15112.8		-16207.6		-16533.5
LL(β)			-11919.4		-11527.9		-11610.2		-12146.6		-11610.5
rho-bar²			0.131		0.190		0.231		0.250		0.298

stochastic sim²		very small var		small var		average var		large var		very large var	
parameter	value	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat
μ	1.00	0.439	20.8	0.590	15.2	0.750	9.3	0.789	7.8	0.849	5.6
β_{bumps}	-0.10	-0.098	0.1	-0.100	0.0	-0.106	-0.3	-0.106	-0.3	-0.111	-0.5
β_{turns}	-0.30	-0.264	2.6	-0.264	2.6	-0.259	2.9	-0.259	2.9	-0.256	3.1
β_{PSC}	1.00	0.851	-3.7	0.846	-3.8	0.818	-4.5	0.799	-5.0	0.744	-6.4
LL(0)			-13818.3		-14299.9		-15278.2		-15637.4		-16284.3
LL(β)			-11765.2		-11444.2		-11447.5		-11064.4		-11043.7
rho-bar²			0.148		0.199		0.250		0.292		0.322

random walk		very small var		small var		average var		large var		very large var	
parameter	value	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat
μ	1.00	0.950	1.9	0.976	0.9	0.977	0.9	0.974	1.0	0.973	1.0
β_{bumps}	-0.10	-0.110	-0.5	-0.114	-0.6	-0.124	-1.1	-0.104	-0.2	-0.109	-0.4
β_{turns}	-0.30	-0.285	1.1	-0.284	1.1	-0.288	0.9	-0.289	0.8	-0.287	0.9
β_{PSC}	1.00	0.938	-1.5	0.942	-1.4	0.940	-1.5	0.939	-1.5	0.943	-1.4
β_{lnkq}	1.00	1.000	-	1.000	-	1.000	-	1.000	-	1.000	-
LL(0)			-14011.3		-14606.8		-15141.3		-15799.5		-16307.9
LL(β)			-11810.5		-11240.6		-11777.5		-11422.2		-10813.9
rho-bar²			0.157		0.230		0.222		0.277		0.337

Notes: β_{length} = -1, 4000 observations, t-statistic with respect to the "true value".

TABLE 6 Meta-Analysis Estimates of the Accuracy of Model Parameters

Parameter	est.	t-stat
characteristic related to the technique used to generate choice sets		
deterministic technique [a]	-	-
stochastic technique	16.197	30.00
k-shortest path [a]	-	-
link penalty	3.544	6.53
branch-and-bound	3.630	6.68
stochastic simulation	-6.971	-18.45
doubly stochastic simulation	-8.580	-19.67
random walk [a]	-	-
small variance	-5.347	-20.48
average variance [a]	-	-
large variance	0.344	1.35
characteristic related to the technique used to postulate choices		
low heterogeneity [a]	-	-
medium/high heterogeneity	-7.810	-18.97
small choice set size	-5.914	-15.71
medium choice set size [a]	-	-
large choice set size	6.484	27.82
consistent with generation	33.748	73.38
constant	10.712	35.56
N		3600
R^2		0.698

Notes: [a] reference category.

Again, the difficulty in interpreting estimation results from table 5 motivates the meta-analysis of the accuracy of parameter estimates for the 900 models. Table 6 presents the regression model over all the 3,600 parameters, while table 7 focuses on each single parameter. Unbiased estimates are obtained when the same path generation technique is applied for objective and subjective choice sets. An increase in accuracy is related to the implementation of stochastic approaches, preferably random walk rather than stochastic simulation. Large variance does not increase estimation accuracy with respect to average variance, and small variance decreases it significantly. Estimation accuracy grows when models are estimated with respect to postulated behavior from large choice sets with high degree of similarity across alternatives.

Similar results are found when regression models consider single parameters. Notable differences are the comparable effectiveness in estimating scale parameters with choice sets generated with different stochastic techniques, the inferior relevance of stochastic techniques in estimating scale parameters, and the superior relevance of random walk in estimating path size estimates. The general interpretation of these differences does not seem intuitive, since different techniques are not expected having different effects on the various estimates.

TABLE 7 Meta-Analysis Estimates of the Accuracy of Single Parameters

	μ		β_{bumps}		β_{turns}		β_{PSC}	
Parameter	est.	t-stat	est.	t-stat	est.	t-stat	est.	t-stat
Characteristic related to the technique used to generate the choice sets								
deterministic technique [a]	-	-	-	-	-	-	-	-
stochastic technique	3.976	4.28	19.375	16.85	22.559	19.90	19.861	20.78
k-shortest path [a]	-	-	-	-	-	-	-	-
link penalty	4.218	4.52	-0.627	-0.54	7.176	6.29	4.310	4.48
branch-and-bound	6.491	6.95	-3.286	-2.84	4.115	3.61	8.102	8.43
stochastic simulation	-1.739	-2.68	-6.262	-7.78	-5.055	-6.37	-15.105	-22.58
doubly stochastic simulation	-1.102	-1.47	-8.667	-9.33	-7.529	-8.22	-17.395	-22.52
random walk [a]	-	-	-	-	-	-	-	-
small variance	-9.866	-21.98	-3.986	-7.17	-5.193	-9.47	-2.659	-5.75
average variance [a]	-	-	-	-	-	-	-	-
large variance	-0.524	-1.19	-0.072	-0.13	0.338	0.63	0.518	1.14
Characteristic related to the technique used to postulate choices								
low heterogeneity [a]	-	-	-	-	-	-	-	-
high heterogeneity	1.568	2.22	-9.401	-10.72	-13.334	-15.43	-10.537	-14.45
small choice set size	-2.069	-3.20	-8.610	-10.74	-8.876	-11.23	-4.493	-6.74
medium choice set size [a]	-	-	-	-	-	-	-	-
large choice set size	4.235	10.57	9.240	18.61	8.018	16.39	4.823	11.69
consistent with generation	31.043	39.26	28.993	29.59	39.628	41.04	33.498	41.14
constant	8.914	17.21	15.598	24.31	11.162	17.65	6.720	12.60
N	900		900		900		900	
R^2	0.626		0.790		0.870		0.648	

Notes: [a] reference category.

4.3. Accuracy of flow predictions

Table 8 summarizes the RMSE and MAPE when estimated models are applied to the dataset used for obtaining the "postulated chosen routes" and "simulated predicted routes" are compared to the "postulated predicted routes" in terms of link flows. Similarly to the coverage, as 900 comparisons are computed, the table summarizes the relationship of the 30 variations of path generation techniques with respect to “postulated chosen routes” by combining results for the five variations of each technique. Expectedly, stochastic techniques significantly outperform deterministic ones, most likely because of the more realistic behavioral assumptions that for example allow generating different choice sets for different travelers. However, link penalty unexpectedly outperforms both k-shortest path and branch-and-bound, even though its behavioral assumption is simpler and estimation results do not suggest better modeling performances. None of the three stochastic techniques emerges as preferable for prediction purposes, an interesting result when considering that the random walk produces better estimates. Most likely, the fact that the correction term is not used for prediction purposes reduces the advantage for the random walk.

TABLE 8 Prediction Errors of Generated Choice Sets with Respect to Postulated Choice Sets

Path generation technique	variation	k-path	link penalty	branch and bound	stochastic simulation	doubly stochastic simulation	random walk
k-shortest path	RMSE	0.5765	0.9234	1.0198	0.4390	0.4136	0.4663
	MAPE	0.0950	0.1105	0.1114	0.1092	0.1052	0.1172
link penalty	RMSE	0.2508	0.1420	0.2213	0.1597	0.1658	0.1882
	MAPE	0.0670	0.0412	0.0569	0.0514	0.0547	0.0622
branch-and-bound	RMSE	0.4654	0.3755	0.2920	0.4158	0.4275	0.4534
	MAPE	0.1195	0.0982	0.0739	0.1201	0.1234	0.1335
stochastic simulation	RMSE	0.2092	0.1476	0.2177	0.1088	0.1019	0.1728
	MAPE	0.0505	0.0401	0.0554	0.0372	0.0356	0.0583
doubly stochastic simulation	RMSE	0.2055	0.1488	0.2212	0.1092	0.1006	0.1731
	MAPE	0.0495	0.0403	0.0560	0.0372	0.0351	0.0584
random walk	RMSE	0.2081	0.1769	0.2340	0.1190	0.1016	0.1837
	MAPE	0.0459	0.0452	0.0571	0.0388	0.0339	0.0594

Again, the difficulty in interpreting prediction results from table 8 suggests the meta-analysis of the MAPE. Table 9 presents the regression model for the 900 comparisons, whose interpretation considers that modeling the error implies negative estimates should be interpreted as positive relation to prediction accuracy. As for the estimation accuracy, an increase in prediction accuracy is related to the implementation of stochastic approaches, preferably random walk rather than stochastic simulation and doubly stochastic simulation. Unlike for the estimation accuracy, enlarging variance helps in prediction, suggesting that modelers should generate large choice sets for traffic assignment in order to better reproduce predicted flows, regardless of the postulated behavior.

TABLE 9 Meta-Analysis Estimates of the Mean Average Prediction Error

Parameter	est.	t-stat
characteristic related to the technique used to generate choice sets		
deterministic technique [a]	-	-
stochastic technique	-0.400	-15.54
k-shortest path [a]	-	-
link penalty	-0.376	-14.52
branch-and-bound	-0.209	-8.06
stochastic simulation	0.107	5.92
doubly stochastic simulation	0.143	6.85

random walk [a]	-	-
small variance	0.123	9.84
average variance [a]	-	-
large variance	-0.035	-2.89
characteristic related to the technique used to postulate choices		
low heterogeneity [a]	-	-
medium/high heterogeneity	0.181	9.20
small choice set size	-0.019	-1.06
medium choice set size [a]	-	-
large choice set size	-0.089	-8.00
consistent with generation	-0.101	-4.60
constant	0.121	25.31
N	900	
R^2	0.835	

Notes: [a] reference category.

5. SUMMARY AND CONCLUSIONS

In the context of modeling route choice behavior, modelers generate objective choice sets by selecting a path generation technique and its parameters according to personal judgments. The literature demonstrates that these personal judgments affect model estimates and predictions (e.g., *21*, *22*, *23*, *24*), but fails to provide suggestions for the implementation of path generation techniques. The current study proposes a methodology and an experimental setting that evaluates the effect of path generation techniques on model estimates and flow predictions and leads to general indications about personal judgments in path generation.

Initially, path generation techniques are implemented within a synthetic network to generate subjective choice sets possibly considered by travelers. Next, "true model estimates" and "postulated predicted routes" are assumed from the simulation of a route choice model. Then, path generation techniques are applied to generate objective choice sets for model estimation and to compare estimates with the postulated "true model estimates". Last, predictions from the postulation with subjective choice sets and from the simulation of models estimated with objective choice sets are compared in terms of link flows.

This paper illustrates the importance of a meta-analytic approach in the synthesis of a large number of models generate a large amount of results that are otherwise difficult to summarize and to process. Summary statistics only partially capture the influence of the characteristics of path generation techniques on model estimates and flow predictions. On the contrary, meta-analysis successfully summarizes the relevance of judgments in the selection of path generation techniques and their parameters for increasing coverage of observed behavior and augmenting accuracy of model estimation.

The approach seems easily transferable to any study concentrating on the estimation of a large number of models and requiring a summary of the results without involving data mining or Bayesian inference that would be much more expensive from a conceptual and a computational perspective (*29*). Moreover, the fit of the meta-analytic models indicates that the variation in the results can indeed be explained by modeling judgments.

The value of the approach consists in providing the first comprehensive indications about judgments in the implementation of path generation techniques following the execution of a systematic and extensive experimental setting. The coverage of chosen routes increases with the implementation of stochastic techniques and the selection of average to large variance in their distribution parameters, with better results obtained when covering behavior postulated from large choice sets containing mainly similar alternatives. The accuracy of model estimation grows with stochastic methods and in particular with the random walk algorithm that exploits the correction for the unequal sampling probabilities of the generated routes. Variance of the parameters and estimation of models with the aforementioned characteristics of large size and high similarity also contribute to estimation accuracy. The accuracy of flow prediction parallels the one of model estimation, with the difference that generating even larger choice sets seems to improve prediction performances.

Results suggest that modelers should apply stochastic approaches with the possibility of correcting for unequal sampling probability while maintaining a fairly reasonable level of variance. Estimation of models would greatly improve and the issue of the coverage of observed behavior would not be raised because the correction would cover for adding alternatives not generated. Only the random walk algorithm currently provides this opportunity, as a correction term exists for the stochastic simulation (*30*), but presents severe shortcomings. Specifically, problems exist in the approximation of the Gumbel distribution with a normal distribution of route costs obtained by summing means and variances over the links and in the assumption of equal variance of the route cost distributions across alternatives.

Further research should address the need for a theoretically sound correction term for both stochastic simulation approaches, a verification of the efficiency of random walk algorithm in more complex networks, and an analysis of the effects of path generation techniques on model estimates for larger networks.

ACKNOWLEDGEMENTS

The author would like to thank Emma Frejinger for providing the details of the Borlange network.

REFERENCES

1. Richardson, A. Search models and choice set generation. *Transportation Research Part A: Policy and Practice*, *Vol. 16*, 1982, pp. 403-419.
2. Horowitz, J.L., and J.J. Louviere. What is the role of consideration sets in choice modeling? *International Journal of Research in Marketing*, *Vol. 12*, 1995, pp. 39-54.
3. Kaplan, S., and C.G. Prato. Joint modeling of constrained path enumeration and path choice behavior: a semi-compensatory approach. *Proceedings of the European Transport Conference*, 2010, Glasgow, Scotland.
4. Bovy, P.H.L. On modelling route choice sets in transportation networks: a synthesis. *Transport Reviews*, *Vol. 29*, 2009, pp. 43-68.
5. Prato, C.G. Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, *Vol. 2*, 2009, pp. 65-100.
6. Cascetta, E., F. Russo, and A. Vitetta. Stochastic user equilibrium assignment with explicit path enumeration: comparison of models and algorithms. In: Papageorgiou, M., and A. Pouliezos (eds.), *Proceedings of 8th International Symposium on Transportation Systems*, pp. 1031-1037, 1997, Elsevier, Oxford, U.K.
7. Bekhor, S., and T. Toledo. Investigating path-based solution algorithms to the stochastic user equilibrium problem. *Transportation Research Part B: Methodological*, *Vol. 39*, 2005, pp. 279-295.

8. Bliemer, M.C.J., and H. Taale. Route generation and dynamic traffic assignment for large networks. *Proceedings of the First International Symposium on Dynamic Traffic Assignment* pp. 90-99, 2006, Leeds, U.K.
9. Lombard, K., and R.L. Church. The gateway shortest path problem: generating alternative routes for a corridor location problem. *Geographical Systems*, *Vol. 1*, 1993, pp. 25-45.
10. Hunt, D.T., and A.L. Kornhauser. Assigning traffic over essentially-least-cost paths. *Transportation Research Record: Journal of the Transportation Research Board*, *Vol. 1556*, 1997, pp. 1-7.
11. Van der Zijpp, N.J., and S. Fiorenzo-Catalano. Path enumeration by finding the constrained K-shortest paths. *Transportation Research Part B: Methodological*, *Vol. 39*, 2005, pp. 545-563.
12. Ben-Akiva, M.E., M.J. Bergman, A.J. Daly, and R. Ramaswamy. Modeling inter-urban route choice behaviour. In: Volmuller, J., and R. Hamerslag (eds.), *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, VNU Science Press, Utrecht, The Netherlands, 1984, pp. 299-330.
13. Azevedo, J.A., M.E.O. Santos Costa, J.J.E.R. Silvestre Madera, and E.Q. Vieira Martins. An algorithm for the ranking of shortest paths. *European Journal of Operational Research*, *Vol. 69*, 1993, pp. 97-106.
14. De la Barra, T., B. Perez, and J. Anez. Multidimensional path search and assignment. *Proceedings of the 21st PTRC Summer Annual Meeting*, pp. 307-319, 1993, Manchester, U.K.
15. Prato, C.G., and S. Bekhor. Applying branch & bound technique to route choice set generation. *Transportation Research Record: Journal of the Transportation Research Board*, *Vol. 1985*, 2006, pp. 19-28.
16. Ramming, S. *Network Knowledge and Route Choice*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2002.
17. Bekhor, S., M.E. Ben-Akiva, and S. Ramming. Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, *Vol. 144*, 2006, pp. 235-247.
18. Nielsen, O.A. A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B: Methodological*, *Vol. 34*, 2000, pp. 377-402.
19. Bovy, P.H.L., and S. Fiorenzo-Catalano. Stochastic route choice set generation: behavioral and probabilistic foundations. *Transportmetrica*, *Vol. 3*, 2007, pp. 173-189.
20. Frejinger, E., M. Bierlaire, and M.E. Ben-Akiva. Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, *Vol. 43*, 2009, pp. 984-994.
21. Bekhor, S., and C.G. Prato. Effects of choice set composition in route choice modeling. *Proceedings of the 11th Conference of the International Association for Travel Behavior Research*, 2006, Kyoto, Japan.
22. Prato, C.G., and S. Bekhor. Modeling route choice behavior: how relevant is the choice set composition? *Transportation Research Record: Journal of the Transportation Research Board*, *Vol. 2003*, 2007, pp. 64-73.

23. Bliemer, M.C.J., and P.H.L. Bovy. Impact of route choice set on route choice probabilities. *Transportation Research Record: Journal of the Transportation Research Board*, *Vol. 2076*, 2008, pp. 10-19.

24. Bekhor, S., T. Toledo, and J.N. Prashker. Implementation issues of route choice models in path-based algorithms. *Transportmetrica*, *Vol. 4*, 2008, pp. 117-133.

25. Sheffi, Y., and W.B. Powell. An algorithm for the equilibrium assignment problem with random link times. *Networks*, *Vol. 12*, 1982, pp. 191-207.

26. Bierlaire, M., and E. Frejinger. Route choice models with subpath components. *Proceedings of the 5th Swiss Transport Research Conference*, 2005, Ascona, Switzerland.

27. Nielsen, O.A., A. Daly, and R.D. Frederiksen. A stochastic multi-class road assignment model with distributed time and cost coefficients. *Networks and Spatial Economics*, *Vol. 2*, 2002, pp. 327-346.

28. Bovy, P.H.L., S. Bekhor, and C.G. Prato. The factor of revised path size: an alternative derivation. *Transportation Research Record: Journal of the Transportation Research Board*, *Vol. 2076*, 2008, pp. 132-140.

29. Leamer, E.E. Let's take the con out of econometrics. *American Economic Review*, *Vol. 73*, 1983, pp. 31-43.

30. Bovy, P.H.L., S. Bekhor, and C.G. Prato. Route sampling correction for stochastic route choice set generation. *Proceedings of the 88th Annual Meeting of the Transportation Research Board*, 2009, Washington, D.C.