

META-ANALYSIS OF CHOICE SET GENERATION EFFECTS ON ROUTE CHOICE MODEL ESTIMATES AND PREDICTIONS

Carlo Giacomo Prato

*Dept of Transport, Technical University of Denmark, Bygningstorvet 116 Vest,
2800 Kgs. Lyngby, Denmark
E-mail: cgp@transport.dtu.dk*

Submitted 23 June 2011; accepted 5 September 2011

Abstract. Large scale applications of behaviorally realistic transport models pose several challenges to transport modelers on both the demand and the supply sides. On the supply side, path-based solutions to the user assignment equilibrium problem help modelers in enhancing the route choice behavior modeling, but require them to generate choice sets by selecting a path generation technique and its parameters according to personal judgments. This paper proposes a methodology and an experimental setting to provide general indications about objective judgments for an 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’. A meta-analytical approach allows synthesizing the effect of judgments for 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 transport modelers should implement stochastic path generation techniques with average variance of its distribution parameters and correction for unequal sampling probabilities of the alternative routes in order to obtain satisfactory results in terms of coverage of ‘postulated chosen routes’, reproduction of ‘true model estimates’ and prediction of ‘postulated predicted routes’.

Keywords: path-based route choice modeling, meta-analysis, path generation, model estimation, model prediction, large scale model applications, path size correction, logit structure.

1. Introduction

In recent years, technological and theoretical enhancements allowed transport planning to move towards large scale applications of behaviorally realistic models. A number of activity-based model systems have been designed for large metropolitan areas such as Portland (Bowman *et al.* 1998), San Francisco (Bradley *et al.* 2001), New York (Vovsha *et al.* 2002), Columbus (Vovsha *et al.* 2004), Atlanta (PBConsult 2004), Sacramento (Bradley *et al.* 2007), Dallas (Pinjari *et al.* 2008), Tel-Aviv (Cambridge Systematics 2008) and Jakarta (Yagi, Mohammadian 2011), and agent-based model systems have been developed for even larger areas such as Switzerland (Meister *et al.* 2010). Large scale applications pose several challenges to transport modelers from the computational perspective, and judgments supported from theoretical and empirical considerations enable transport modelers to search for effective estimation and running procedures.

While travel demand representation has flourished with the development of micro-simulation activity-based model frameworks that are theoretically advanced and behaviorally realistic, travel supply representation for large scale networks has not equally thrived. Path-based solutions to the user equilibrium assignment problem have the potential to help transport modelers in enhancing route choice models in static or dynamic large scale applications.

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 that call for separate modeling: choice set formation is trial-and-error determined (Richardson 1982), preference-driven (Horowitz, Louviere 1995) and constraint-related (Kaplan, Prato 2010), while choice from alternatives is usually represented as a compensatory decision (see,

e.g., Bovy 2009; Prato 2009). 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 (Cascetta *et al.* 1997), unrealistic and inefficient paths are found within implicit choice sets for link-based assignment (Bekhor, Toledo 2005), and theoretical and computational advantages are shown when the choice set generation is separately performed prior to traffic assignment (Bliemer, Taale 2006).

Several solutions have been proposed to the explicit path generation problem. Deterministic solutions include variations of shortest path algorithms (e.g., Hunt, Kornhauser 1996; Lombard, Church 1993; Van der Zijpp, Fiorenzo-Catalano 2005), minimization of generalized cost functions (Ben-Akiva *et al.* 1984), application of heuristic rules combined with the shortest path searches (e.g., Azevedo *et al.* 1993; De la Barra *et al.* 1993), and implementation of a branch-and-bound algorithm (Prato, Bekhor 2006). Stochastic solutions include single stochastic simulation (e.g., Ramming 2002; Bekhor *et al.* 2006), doubly stochastic simulation (Nielsen 2000; Bovy, Fiorenzo-Catalano 2007) and a random walk algorithm (Frejinger *et al.* 2009). Advantages and disadvantages related to the implementation of existing path generation techniques are extensively discussed by Bovy (2009) and Prato (2009).

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 transport 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 role of path generation techniques on model estimates and flow predictions has not been documented. Model performances of several route choice models estimated with different choice sets have been compared in terms of likelihood values (Bekhor, Prato 2006; Prato, Bekhor 2007), but 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 modes estimated for a small synthetic network have been paralleled to choice probabilities from a postulated probit model (Bliemer, Bovy 2008), but the analysis fails to assess the effects of the implementation of path generation techniques because of the peculiar context with a universal realm of only 12 alternatives. The influence of choice set size on objective function values and convergence times of solutions to the SUE problem have been examined (Bekhor *et al.* 2008), but the comparison fails to investigate path generation techniques other than the *k*-shortest path algorithm. In a nutshell, existing studies about the choice set effects on route choice models focus on the analysis of the robustness of models and methods, rather than on actual effects of path generation

techniques and on the provision of general indications about judgments for generating objective choice sets.

This study presents the first systematic analysis of the judgments that the transport modelers are required to apply efficient and unbiased path generation, and proposes guidelines for effective implementation of path generation techniques.

The analysis implements several path generation techniques to a synthetic network. Initially, choice sets generated with different techniques are used for defining possible subjective choice sets for the postulation of 'true model estimates' of a route choice model and the simulation of 'postulated chosen routes'. Next, choice sets generated with different techniques are employed for constructing objective choice sets for model estimation. Then, route choice models are estimated for all the possible combinations of subjective and objective choice sets. Last, predictions from the simulation of models estimated with objective choice sets are compared to the 'postulated chosen routes'. The advantage of this approach is three-fold:

- 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;
- estimating the same model specification within the same synthetic network allows isolating choice set effects from model and network effects;
- 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-analytical 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 within the same study 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 guidelines 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 parameters to be implemented are examined by considering a variety of techniques that are included in the analysis according to the rationale presented in four points.

The first point 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 point involves the selection of deterministic approaches. Even though their evolution suggests the superiority of more recent developments with respect to the shortest path algorithms, considering the only one deterministic technique would bias again the current analysis. The first deterministic technique is considered to be the largely applied and the most straightforward approach to the choice set generation problem consisting in the computation of k -shortest paths. The second deterministic technique is considered to be the iteration of the shortest path search after heuristic rules which penalize links on the last shortest path computed in the iterative process (De la Barra *et al.* 1993), which is preferable to link elimination (Azevedo *et al.* 1993) that introduces network disconnection problems. The third deterministic technique is considered to be the enumeration of the paths connecting the origin and destination of a trip under behavioral and logical constraints within a branch-and-bound algorithm (Prato, Bekhor 2006).

The third point concerns the selection of stochastic approaches. The current analysis examines the effects of three techniques most recently developed and used in the literature about 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., Bekhor *et al.* 2006; Bovy, Fiorenzo-Catalano 2007). The second stochastic technique considered is the natural evolution of the previous approach considering an error term for traffic network variations and an error term for traveler taste heterogeneity (e.g., Nielsen 2000; Bovy, Fiorenzo-Catalano 2007).

The third stochastic technique considered is a random walk algorithm that is biased towards the search for the shortest path (Frejinger *et al.* 2009).

The fourth point playing a role in the selection of path generation techniques involves the parameters for their implementation being relevant for model estimation and model implementation. Considering only one set of parameters (e.g., number of iterations, specific parameters, probability distributions) it would bias the current analysis that intends to provide general guidelines about judgments concerning the selection of path generation techniques.

With respect to the k -shortest path, five values of k cover from a fairly small to a very large window of admissible path costs. With respect to the link penalty,

five combinations with increasing penalizing factor cover from a small to a large variation of the alternatives in the generated choice sets. With respect to the branch-and-bound, five combinations of the thresholds of the branching rule allow assessing the effect of increasing choice set size and route heterogeneity in the generation process.

With respect to the stochastic simulation, as unfortunately normal distribution produces the negative link impedances and truncated normal distribution is not additive over the links, five combinations of shape and scale of a gamma distribution with mean equal to the link impedance and a range of increasing variances assure a large variety of positive link impedances. With respect to the doubly stochastic simulation, five variations of a gamma distribution for the first error component representing travel time variation and a log-normal distribution for the second error component capturing taste heterogeneity cover a range of alternatives with growing variances. With respect to the random walk algorithm, five values of the parameters of the Kumaraswami distribution of the weights encompass a range of variance with respect to the shortest path in the generation process.

In summary, five variations are considered for each of the six path generation techniques selected 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 path generation techniques to the synthetic network represented in Figure that consists of 38 nodes and 64 links, with link length proportional to the length of the figure and some links with 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.* (2009) for testing the random walk algorithm. The universal realm consists of 170 alternative routes between origin O and destination D , among which 29 have equal minimum length.

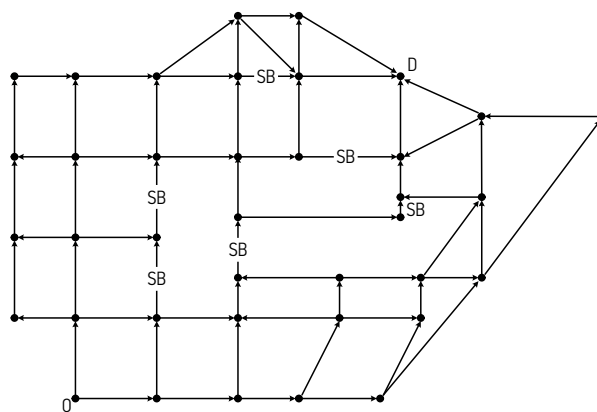


Fig. Network for experimental design (Frejinger *et al.* 2009)

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

Datasets of 4000 observations for estimation purposes and 1000 observations for prediction purposes are generated from the variations of path generation techniques applied. This procedure creates 30 datasets of 4000 subjective choice sets for model estimation and 30 datasets of 1000 subjective choice sets for model prediction.

For each dataset of subjective choice sets, a PSC-Logit model (Bovy *et al.* 2008) is postulated. The advantage in using the PSC-Logit is two-fold:

- the model accounts for similarities across alternatives while maintaining a simple Logit structure;
- MNL modifications are robust with respect to variations in the number of alternatives and in the composition of the choice sets (e.g., Bliemer, Bovy 2008; Prato, Bekhor 2007).

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

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

where: $Length_j$ is the length; $SpeedBumps_j$ is the number of speed bumps; $Turns_j$ is the number of turns; 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 (Bovy *et al.* 2008, 2009):

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

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 ; 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, 4000 ‘postulated chosen routes’ for estimation purposes and 1000 ‘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. Implementation of path generation techniques

| Path generation technique | parameter | very small variance | small variance | average variance | large variance | very large variance |
|--|----------------|---------------------|----------------|------------------|----------------|---------------------|
| define length limit in order to define k -shortest paths | | | | | | |
| k -shortest path | length limit | 35 | 36 | 37 | 40 | 51 |
| | k | 33 | 51 | 88 | 104 | 170 |
| Iterate the shortest path searches and penalize the shortest path links | | | | | | |
| link penalty | penalty factor | 2% | 3% | 5% | 10% | 20% |
| | iterations | 50 | 50 | 50 | 100 | 100 |
| connect origin and destination of the trips with five behavioral and logical thresholds | | | | | | |
| branch-and-bound | 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 |
| iterate the shortest path searches after extracting link length from gamma distribution | | | | | | |
| stochastic simulation | mean | length | length | length | length | length |
| | st.dev | 0.25 length | 0.50 length | length | 2 length | 3 length |
| iterate the shortest path searches after extracting link length from gamma distribution and travelers’ preferences from log-normal distribution | | | | | | |
| doubly stochastic simulation | mean | length | length | length | length | length |
| | st.dev | 0.50 length | 0.50 length | length | length | 2 length |
| | mean | -1 | -1 | -1 | -1 | -1 |
| | st.dev | 0.25 | 0.25 | 0.5 | 1 | 1 |
| calculate route probabilities from link probabilities based on link weights that are Kumaraswami distributed with two parameters b_1 and b_2 | | | | | | |
| random walk | 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 same 30 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:

$$Cov_{obj-sub} = \frac{\sum I(R_{sub,n} \in C_{obj,n})}{N} \cdot 100, \quad (3)$$

where: $Cov_{obj-sub}$ 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 utility function:

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

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 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 utility function is:

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

where: k_{jn} is the number of time route j is sampled for observation n ; q_j is the probability of sampling route j ; the additional 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 3600 estimated parameters:

$$Acc_{par,obj-sub} = \text{Prob} \left(\frac{est_{par,obj-sub} - exp_{par}}{stderrest_{par,obj-sub}} < t \right) \cdot 100, \quad (6)$$

where: $Acc_{par,obj-sub}$ 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; $est_{par,obj-sub}$ is the estimate; $stderrest_{par,obj-sub}$ is its standard error; exp_{par} is the expected ‘true model estimate’; 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 from the estimates of the 30 models using the same dataset for obtaining the ‘postulated chosen routes’. 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. Predictions from the estimated models are evaluated with the calculation of the following error measures for each combination of estimated model with the respective dataset of subjective choice sets:

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

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

where: $RMSE_{sim-pos}$ is the root mean square error; $MAPE_{sim-pos}$ 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’; $N_{pos,a}$ is the flow on link a as calculated by translating the ‘postulated predicted routes’.

Given the 900 coverage values, the 3600 accuracy values from model estimation, and the 900 mean absolute percentage errors, a 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. Expectedly:

- the increase in the variance of the parameters for each path generation technique produces larger choice sets;
- deterministic techniques generate the same alternative routes for the same origin-destination pair, an unreasonable outcome when considering that most likely different travelers have different subjective choice sets;
- stochastic techniques produce different alternative routes for the same origin-destination pair, a behaviorally plausible trait.

Table 3 summarizes the coverage of the dataset of objective choice sets with respect to the ‘postulated cho-

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 |
|------------------------------|---------|---------------------|----------------|------------------|----------------|---------------------|
| <i>k</i> -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. Coverage of objective choice sets with respect to ‘postulated chosen routes’

| Path generation technique | variation | <i>k</i> -path | link penalty | branch and bound | stochastic simulation | doubly stochastic simulation | random walk |
|------------------------------|----------------|----------------|--------------|------------------|-----------------------|------------------------------|-------------|
| <i>k</i> -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 |

sen 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. Expected findings are found from this snapshot, regardless of the postulated behavior:

- enlarging the number of alternatives considered in the k -shortest path and relaxing the thresholds in the branch-and-bound the algorithm increases in coverage;
- among deterministic techniques, the branch-and-bound outperforms competing techniques;
- among stochastic techniques, doubly stochastic simulation outperforms competing techniques.

The most unexpected finding from this snapshot is that increasing the variance and enlarging objective choice sets does not boost the coverage. Possibly, the snapshot is limited because 'postulated predicted routes' consider jointly all five variations for each technique and lose part of the information. Also, larger variance of the parameters of path generation techniques most likely generates irrelevant routes that are not created under different parameters.

The difficulty in interpreting coverage results from Table 3 motivates the meta-analysis for the 900 combinations between datasets of objective choice sets and 'postulated chosen routes'. 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 the 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 term 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 a 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. Lastly, the parameter β_{bumps} seems to be recovered almost consistently, while the scale parameter μ appears to be recovered only sporadically.

Table 5. Example of model estimates for generated choice sets from different path generation techniques

| <i>k</i> -shortest path | | very small var | | small var | | average var | | large var | | very large var | |
|-----------------------------|-------|----------------|----------------|-----------|----------------|-------------|----------------|-----------|----------------|----------------|----------------|
| Parameter | value | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -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 | |
| $LL(\beta)$ | | -7284.0 | | -10369.2 | | -12311.0 | | -13471.8 | | -13472.3 | |
| $\rho\text{-bar}^2$ | | 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. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -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 | |
| $LL(\beta)$ | | -12688.1 | | -12244.1 | | -12021.2 | | -12142.4 | | -11591.2 | |
| $\rho\text{-bar}^2$ | | 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. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -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 | |
| $LL(\beta)$ | | -8997.5 | | -10413.5 | | -10575.6 | | -11303.3 | | -11863.4 | |
| $\rho\text{-bar}^2$ | | 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. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -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 | |
| $LL(\beta)$ | | -11919.4 | | -11527.9 | | -11610.2 | | -12146.6 | | -11610.5 | |
| $\rho\text{-bar}^2$ | | 0.131 | | 0.190 | | 0.231 | | 0.250 | | 0.298 | |
| stochastic sim ² | | very small var | | small var | | average var | | large var | | very large var | |
| Parameter | value | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -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 | |
| $LL(\beta)$ | | -11765.2 | | -11444.2 | | -11447.5 | | -11064.4 | | -11043.7 | |
| $\rho\text{-bar}^2$ | | 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. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -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_{linkq} | 1.00 | 1.000 | - | 1.000 | - | 1.000 | - | 1.000 | - | 1.000 | - |
| $LL(0)$ | | -14011.3 | | -14606.8 | | -15141.3 | | -15799.5 | | -16307.9 | |
| $LL(\beta)$ | | -11810.5 | | -11240.6 | | -11777.5 | | -11422.2 | | -10813.9 | |
| $\rho\text{-bar}^2$ | | 0.157 | | 0.230 | | 0.222 | | 0.277 | | 0.337 | |

Notes: $\beta_{length} = -14000$ observations; *t*-statistic with respect to the ‘true value’

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 3600 parameters, while Table 7 focuses on each single parameter. Unbiased estimates require considering the first and the foremost which is the same path generation technique for objective and subjective choice sets and that obviously increases its estimation accuracy the most. 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 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: ^areference 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 five variations of each technique. Expectedly, stochastic techniques outperform deterministic ones significantly, most likely because of better behavioral assumptions that for example allow generating different choice sets for different travelers. Less predictably, link penalty 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 for prediction purposes the correction term is not used reduces the advantage for the random walk ability of predicting correct routes.

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 should consider that modeling the error implies negative estimates which 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 reproduce predicted flows better, regardless of the postulated behavior.

Obviously, correct predictions are influenced by correct reproduction of the ‘true values’ of the model estimates. Thus, conclusions about the effects of path generation techniques on model predictions are more complex to discern and to generalize because results are dependent on the estimation accuracy.

5. Summary and Conclusions

Path-based solutions to the user equilibrium assignment problem help transport modelers in enhancing the route choice models in large scale applications. The implementation of path-based solution requires modelers to 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., Bekhor, Prato 2006; Bekhor *et al.* 2008; Bliemer, Bovy 2008; Prato, Bekhor 2007), but fails to suggest guidelines for the implementation of path generation techniques. This paper provides the first comprehensive guidelines about judgments in the implementation of

Table 7. Meta-analysis estimates of the accuracy of single parameters

| Parameter | μ | | β_{bumps} | | β_{turns} | | β_{PSC} | |
|--|--------|----------------|-----------------|----------------|-----------------|----------------|---------------|----------------|
| | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -stat | est. | <i>t</i> -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 |
| <i>k</i> -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 the 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 |
| <i>N</i> | 900 | | 900 | | 900 | | 900 | |
| <i>R</i> ² | 0.626 | | 0.790 | | 0.870 | | 0.648 | |

Notes: ^areference category

Table 8. Prediction errors of generated choice sets with respect to postulated choice sets

| Path generation technique | variation | <i>k</i> -path | link penalty | branch and bound | stochastic simulation | doubly stochastic simulation | random walk |
|------------------------------|-----------|----------------|--------------|------------------|-----------------------|------------------------------|-------------|
| <i>k</i> -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 |

path generation techniques by proposing a methodology and an experimental setting that evaluate the effect of path generation techniques on model estimates and flow predictions.

Initially, path generation techniques are implemented to generate the 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, path generation techniques are applied to generate objective choice sets for model estimation and estimates are compared to the postulated ‘true model estimates.’ Last, predictions from the simulation of models estimated with objective choice sets are compared to the ‘postulated predicted routes.’

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: ^areference category

Results that provide guidelines for efficient path generation by evaluating three requirements for route choice sets:

- the generation of plausible routes that travelers would consider;
- the estimation of route choice models that accurately represent the behavior of travelers;
- the prediction of traffic flows that accurately and consistently measure the network performances in terms of level-of-service.

With respect to the generation of plausible routes that travelers would consider, experimental results suggest that the coverage of observed routes increases with the implementation of stochastic techniques and the selection of average to large variance in their distribution parameters. Moreover, better results are obtained when covering behavior postulated from large choice sets containing mainly similar alternatives.

Because of the accurate estimation of route choice models, experimental results show that the accuracy of model estimates grows with stochastic methods and in particular with the random walk algorithm that corrects 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 increase estimation accuracy.

With respect to the consistent prediction of network performance, experimental results illustrate that the accuracy of flow prediction parallels one of the model estimation, with the difference that generating even larger choice sets seems to improve prediction performances.

The extension of the findings from the analyzed network to large scale applications concerns only computational issues that memory and multiple-core implementation are currently able to resolve. Experimental results suggest that transport modelers should generate routes by applying stochastic approaches with the possibility of correcting for unequal sampling probability while maintaining a reasonable level of variance and generating a large number of routes. Estimation of models would greatly improve and the issue of the coverage of observed behavior would not be raised because the correction would allow accounting for the addition of alternatives not generated. On the one hand, results suggest that transport modelers would greatly benefit from the implementation of the random walk algorithm, since this is the only algorithm that currently provides this opportunity. On the other hand, transport modelers would greatly benefit from the implementation of doubly stochastic path generation techniques that are simple to implement and also for large scale networks, as illustrated for example for the Greater Copenhagen Area (Larsen *et al.* 2010).

In addition, this paper shows the importance of a meta-analytical approach in the synthesis of a large number of models generating 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 (Leamer 1983). It should be noted that the fit of the meta-analytical models indicates that the variation in the results can indeed be explained by modeling judgments.

Further research should address the need for a correction term for unequal sampling probabilities of alternative routes when both stochastic simulation approaches are applied, and the need for the experimental evaluation of the performance of traffic assignment procedures on large scale networks providing that stochastic path generation techniques are implemented.

Acknowledgements

The author would like to thank Emma Frejinger for providing the details of the Borlange network and three anonymous reviewers for the significant contribution in improving a previous version of the manuscript.

References

- Azevedo, J. A.; Santos Costa, M. E. O.; Silvestre Madeira, J. J. E. R.; Vieira Martins, E. Q. 1993. An algorithm for the ranking of shortest paths, *European Journal of Operational Research* 69(1): 97–106. [http://dx.doi.org/10.1016/0377-2217\(93\)90095-5](http://dx.doi.org/10.1016/0377-2217(93)90095-5)
- Bekhor, S.; Ben-Akiva, M. E.; Ramming, S. 2006. Evaluation of choice set generation algorithms for route choice models, *Annals of Operations Research* 144(1): 235–247. <http://dx.doi.org/10.1007/s10479-006-0009-8>
- Bekhor, S.; Prato, C. G. 2006. Effects of choice set composition in route choice modeling, in *Proceedings of the 11th Conference of the International Association for Travel Behavior Research*, 16–20 August, 2006. Kyoto, Japan.
- Bekhor, S.; Toledo, T. 2005. Investigating path-based solution algorithms to the stochastic user equilibrium problem, *Transportation Research Part B: Methodological* 39(3): 279–295. <http://dx.doi.org/10.1016/j.trb.2004.04.006>
- Bekhor, S.; Toledo, T.; Prashker, J. N. 2008. Effects of choice set size and route choice models on path-based traffic assignment, *Transportmetrica* 4(2): 117–133. <http://dx.doi.org/10.1080/18128600808685682>
- Ben-Akiva, M.; Bergman, M. J.; Daly, A. J.; Ramaswamy, R. 1984. Modeling inter-urban route choice behaviour, in *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*. 11–13 July 1984. Delft, The Netherlands, 299–330.
- Bliemer, M. C. J.; Bovy, P. H. L. 2008. Impact of route choice set on route choice probabilities, *Transportation Research Record* 2076: 10–19. <http://dx.doi.org/10.3141/2076-02>
- Bliemer, M. C. J.; Taale, H. 2006. Route generation and dynamic traffic assignment for large networks, in *Proceedings of the First International Symposium on Dynamic Traffic Assignment*, Leeds, U.K., 90–99.
- Bovy, P. H. L. 2009. On modelling route choice sets in transportation networks: a synthesis, *Transport Reviews* 29(1): 43–68. <http://dx.doi.org/10.1080/01441640802078673>
- Bovy, P. H. L.; Bekhor, S.; Prato, C. G. 2008. The factor of revisited path size: alternative derivation, *Transportation Research Record* 2076: 132–140. <http://dx.doi.org/10.3141/2076-15>
- Bovy, P. H. L.; Bekhor, S.; Prato, C. G. 2009. Route sampling correction for stochastic route choice set generation, in *Proceedings of the 88th Annual Meeting of the Transportation Research Board*, 11–15 January 2009, Washington, D.C. 18 p. [DVD].
- Bovy, P. H. L.; Fiorenzo-Catalano, S. 2007. Stochastic route choice set generation: behavioral and probabilistic foundations, *Transportmetrica* 3(3): 173–189. <http://dx.doi.org/10.1080/18128600708685672>
- Bowman, J. L.; Bradley, M.; Shiftan, Y.; Lawton, T. K.; Ben-Akiva, M. 1998. Demonstration of an activity based model system for Portland, in *Proceedings of the 8th World Conference on Transport Research*, 12–17 July 1998, Antwerp, Belgium.
- Bradley, M. A.; Bowman, J. L.; Griesenbeck, B. 2007. Development and application of the SACSIM activity-based model system, in *Proceedings of the 11th World Conference on Transport Research*. 24–26 June 2007, Berkeley, California. 29 p. [CD].
- Bradley, M. A.; Outwater, M. L.; Jonnalagadda, N.; Ruiter, E. 2001. Estimation of an activity-based micro-simulation model for San Francisco, in *Proceedings of the 80th Annual Meeting of the Transportation Research Board*, 7–10 January 2001. Washington, D.C.
- Cambridge Systematics. 2008. *Tel-Aviv Activity Schedule Travel Demand Model System: a Tour-Based Approach*. Report prepared for the Israel Ministry of Transport and Road Safety.
- Cascetta, E.; Russo, F.; Vitetta, A. 1997. Stochastic user equilibrium assignment with explicit path enumeration: comparison of models and algorithms, in *Transportation Systems 1997 (TS'97): a Proceedings Volume from the 8th IFAC/IFIP/IFORS Symposium*, 16–18 June 1997, Chania, Greece, 1031–1037.
- De la Barra, T.; Perez, B.; Anez, J. 1993. Multidimensional path search and assignment, in *Proceedings of the 21st PTRC Summer Annual Meeting*, Manchester, U.K., 307–319.
- Frejinger, E.; Bierlaire, M.; Ben-Akiva, M. E. 2009. Sampling of alternatives for route choice modeling, *Transportation Research Part B: Methodological* 43(10): 984–994. <http://dx.doi.org/10.1016/j.trb.2009.03.001>
- Horowitz, J. L.; Louviere, J. J. 1995. What is the role of consideration sets in choice modeling?, *International Journal of Research in Marketing* 12(1): 39–54. [http://dx.doi.org/10.1016/0167-8116\(95\)00004-L](http://dx.doi.org/10.1016/0167-8116(95)00004-L)
- Hunt, D. T.; Kornhauser, A. L. 1996. Assigning traffic over essentially-least-cost paths, *Transportation Research Record* 1556: 1–7. <http://dx.doi.org/10.3141/1556-01>
- Kaplan, S.; Prato, C. G. 2010. Joint modeling of constrained path enumeration and path choice behavior: a semi-compensatory approach, in *Proceedings of the European Transport Conference*, 11–13 October 2010, Glasgow, Scotland. 15 p. Available from Internet: <http://etcproceedings.org/paper/joint-modeling-of-constrained-path-enumeration-and-path-choice-behavior-a-semi>
- Larsen, M.; Nielsen, O.; Prato, C. G. 2010. Generation and quality assessment of route choice sets in public transport networks by means of RP data analysis, in *Proceedings of the European Transport Conference*, 11–13 October 2010, Glasgow, Scotland.
- Leamer, E. E. 1983. Let's take the con out of econometrics, *The American Economic Review* 73(1): 31–43.
- Lombard, K.; Church, R. L. 1993. The gateway shortest path problem: generating alternative routes for a corridor location problem, *Geographical Systems* 1: 25–45.
- Meister, K.; Balmer, M.; Ciari, F.; Horni, A.; Rieser, M.; Wairach, R. A.; Axhausen, K. W. 2010. Large-scale agent-based travel demand optimization applied to Switzerland, including mode choice, in *Proceedings of the 12th World Conference on Transportation Research*, 16 July 2010, Lisbon, Portugal. <http://dx.doi.org/10.3929/ethz-a-006100288>
- Nielsen, O. A. 2000. A stochastic transit assignment model considering differences in passengers utility functions, *Transportation Research Part B: Methodological* 34(5): 377–402. [http://dx.doi.org/10.1016/S0191-2615\(99\)00029-6](http://dx.doi.org/10.1016/S0191-2615(99)00029-6)
- PBConsult. 2004. *Progress Report for the Year 2003: Modeling Task 2 – Activity/Tour-Based Models*. Regional Transportation Plan Major Update Project for the Atlanta Regional Commission, January, 2004.
- Pinjari, A.; Eluru, N.; Srinivasan, S.; Guo, J. Y.; Sener, I. N.; Bhat, C. R. 2008. CEMDAP: Modeling and microsimulation frameworks, software development, and verification, in *TRB 87th Annual Meeting Compendium of Papers DVD*, 13–17 January 2008, Washington, D.C. [DVD].
- Prato, C. G. 2009. Route choice modeling: past, present and future research directions, *Journal of Choice Modelling* 2(1): 65–100.

- Prato, C. G.; Bekhor, S. 2006. Applying branch-and-bound technique to route choice set generation, *Transportation Research Record* 1985: 19–28.
<http://dx.doi.org/10.3141/1985-03>
- Prato, C. G.; Bekhor, S. 2007. Modeling route choice behavior: how relevant is the composition of choice set?, *Transportation Research Record* 2003: 64–73.
<http://dx.doi.org/10.3141/2003-09>
- Ramming, S. M. 2002. *Network Knowledge and Route Choice*. Submitted to the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Transportation at the Massachusetts Institute of Technology. Available from Internet: <http://dspace.mit.edu/handle/1721.1/49797>
- Richardson, A. 1982. Search models and choice set generation, *Transportation Research Part A: General* 16(5–6): 403–419.
[http://dx.doi.org/10.1016/0191-2607\(82\)90068-1](http://dx.doi.org/10.1016/0191-2607(82)90068-1)
- Van der Zijpp, N. J.; Fiorenzo-Catalano, S. 2005. Path enumeration by finding the constrained K-shortest paths, *Transportation Research Part B: Methodological* 39(6): 545–563.
<http://dx.doi.org/10.1016/j.trb.2004.07.004>
- Vovsha, P.; Petersen, E.; Donnelly, R. 2002. Microsimulation in travel demand modeling: lessons learned from the New York best practice model, *Transportation Research Record* 1805: 68–77. <http://dx.doi.org/10.3141/1805-09>
- Vovsha, P.; Petersen, E.; Donnelly, R. 2004. Impact of intra-household interactions on individual daily activity-travel patterns, *Transportation Research Record* 1898: 87–97.
<http://dx.doi.org/10.3141/1898-11>
- Yagi, S.; Mohammadian, A. 2011. An activity-based microsimulation model of travel demand in the Jakarta metropolitan area, *Journal of Choice Modelling* 3(1): 32–57.