

## Route choice modeling: past, present and future research directions

Carlo Giacomo Prato\*

Transportation Research Institute, Technion - Israel Institute of Technology, Technion City,  
Haifa 32000, Israel

Received 25 September 2007, revised version received 22 January 2009, accepted 8 March 2009

### Abstract

Modeling route choice behavior is problematic, but essential to appraise travelers' perceptions of route characteristics, to forecast travelers' behavior under hypothetical scenarios, to predict future traffic conditions on transportation networks and to understand travelers' reaction and adaptation to sources of information.

This paper reviews the state of the art in the analysis of route choice behavior within the discrete choice modeling framework. The review covers both choice set generation and choice process, since present research directions show growing interest in understanding the role of choice set size and composition on model estimation and flow prediction, while past research directions illustrate larger efforts toward the enhancement of stochastic route choice models rather than toward the development of realistic choice set generation methods.

This paper also envisions future research directions toward the improvement in amount and quality of collected data, the consideration of the latent nature of the set of alternatives, the definition of route relevance and choice set efficiency measures, the specification of models able to contextually account for taste heterogeneity and substitution patterns, and the adoption of random constraint approaches to represent jointly choice set formation and choice process.

*Keywords:* route choice behavior, choice set generation, discrete choice models, substitution patterns

---

\* T: +972-4-8295134, F: +972-4-8225716, [prato@technion.ac.il](mailto:prato@technion.ac.il)



## 1 Introduction

Modeling route choice behavior is essential, given the possibility to appraise travelers' perceptions of route characteristics, to forecast travelers' behavior under hypothetical scenarios, to predict future traffic conditions on transportation networks and to understand travelers' reaction and adaptation to sources of information. Modeling route choice behavior is also problematic, given the complexity of representing human behavior, the lack of travelers' knowledge about the network composition, the uncertainty about travelers' perceptions of route characteristics and the unavailability of exact information about travelers' preferences.

Route choice models not only help analyzing and understanding travelers' behavior, but also constitute the essential part of traffic assignment methods. In the deterministic user equilibrium (DUE) problem, a simple route choice model assumes unrealistically that travelers have perfect knowledge about path costs and choose the route that minimizes their travel costs. In the stochastic user equilibrium (SUE) problem, a probabilistic route choice model assumes reasonably that travelers have imperfect information about path costs and choose the route that minimizes their perceived travel costs given a set of routes. In dynamic traffic assignment (DTA), a route choice model is either predetermined or computed while the network loading mechanism operates.

This paper reviews the state of the art in the analysis of route choice behavior within the discrete choice modeling framework. This review focuses on drivers' route choice behavior in transportation networks, but the same modeling framework is applicable to the multi-modal context (see, e.g., Hoogendoorn-Lanser, 2005). While focusing on the discrete choice modeling framework, this paper does not discuss alternative approaches based on fuzzy logic (e.g., Lotan and Koutsopoulos, 1993; Lotan, 1997; Henn, 2000; Rilett and Park, 2001; Ridwan, 2004), artificial neural networks (e.g., Yang et al., 1993; Dougherty, 1995; Yamamoto et al., 2002) and cognitive psychology (e.g., Nakayama and Kitamura, 2000; Nakayama et al., 2001).

This review examines both major challenges in route choice modeling, namely the generation of a choice set of alternative routes and the estimation of discrete choice models. The first challenge is very demanding, given the lack of detailed information about the actual choice, the scarceness of knowledge about the alternatives actually considered, and the intention of generating heterogeneous and realistic choice sets. The second challenge is also demanding, given the large number of possible alternatives with respect to other discrete choice applications, the high level of correlation among routes that usually share a large number of links, and the interest in understanding how drivers perceive route characteristics and express their preferences. Present and past research directions motivate the review of both choice set generation and discrete choice modeling, since present research directions show growing interest in understanding the role of choice set size and composition on model estimates and traffic flow predictions (Bekhor et al., 2006; Bliemer et al., 2007; Prato and Bekhor, 2007; Bliemer and Bovy, 2008), while past research directions illustrate larger efforts toward the enhancement of stochastic route choice models rather than toward the development of choice set generation methods (Prashker and Bekhor, 2004).

This paper also envisions future research directions. From the data perspective, the amount and the quality of data collected could benefit from an innovative network-free approach to data processing (Bierlaire and Frejinger, 2008). From the choice set generation perspective, the choice set formation process and the latent nature of the set

of alternatives could be considered within stochastic choice set generation models, and relevance and efficiency measures could be defined to evaluate the reliability of generated choice sets. From the discrete choice modeling perspective, a combined error-components and random-parameters formulation could allow to account contemporarily for substitution patterns and taste heterogeneity, and random constraint approaches could jointly represent choice set formation and choice process.

The remainder of the paper is organized as follows. Section 2 introduces the problem of choice set generation by describing the differences between route choice and other travel choices, the purposes for explicit path generation and the requirements for generated choice sets. Then, solutions to the path generation problem are illustrated and their advantages and disadvantages are discussed in light of choice set requirements and existing network applications. Section 3 presents the problem of modeling route choice by illustrating the issue of the large number of alternatives, the matter of the physical overlap among routes and the requirements for route choice models. Then, adaptations of discrete choice models to the route choice context are shown and their benefits and shortcomings are discussed in light of choice set requirements and existing case studies. Section 4 envisions future research directions. Lessons from the past about behavioral assumptions and motivations for developing new methods and models suggest directions for the future in the area of data collection, choice set generation and discrete choice models. Section 5 summarizes major advantages and disadvantages of existing methods and models in the analysis of route choice behavior.

## **2 Choice set generation**

### **2.1 The problem**

Representing route choice behavior consists in modeling the choice of a certain route within a set of alternative routes. This simple definition is not different from other travel choice definitions. In fact, inserting the words “mode” or “destination” in place of the word “route” effectively describes the representation of mode or destination choice behavior.

This semantic similarity hides substantial differences between route choice and other travel choices. In mode choice, a small number of alternatives are easy to define and to visualize (e.g., car, bus, train, bicycle, walking, etc). In destination choice, a large number of alternatives are relatively easy to identify and to envision (e.g., regions, cities, traffic analysis zones, apartments, etc.). In route choice, a vast number of alternative routes are difficult to enumerate and even to visualize. In fact, the population of available routes for a trip in dense urban networks usually consists of hundreds of routes. Moreover, these routes are mostly unknown and “hidden” in the network from which they need to be explicitly extracted, as they are not readily identifiable because of the large number of common links (see Bovy, 2009). The explicit extraction of the routes from the network requires an explicit path generation process.

Accordingly, representing route choice behavior consists in the definition of available alternative routes and the choice of a certain route from the generated choice set. Whether implicit or explicit path generation is preferable constitutes the first aspect of the choice set generation problem considered in this review. A conceptual

rationale and an empirical motive suggest that explicit path generation prior to either model estimation or prediction calculation is preferable.

The conceptual rationale contemplates that choice set formation and choice from considered options are distinct mental processes that require separate modeling, as the first is usually non-compensatory and the second is largely compensatory. Choice set formation is constraint-related, as non-desirable characteristics may lead to route elimination, preference-driven, as the most important choice factors may influence route consideration (e.g., Horowitz and Louviere, 1995), and trial-and-error determined, as a process of route use and information acquisition may affect route perception (e.g., Richardson, 1982). Choice from considered options is largely compensatory, as a trade-off is made among the attributes of interest to maximize travelers' utility. Moreover, explicit choice set generation is especially needed in route choice modeling because of additional reasons, such as the size of the networks, the difference of choice sets between observations and the actual benefits from a policy impact perspective (Bovy, 2009).

The empirical motive considers case studies illustrating the benefits of explicit choice set generation. Cascetta et al. (1997) empirically demonstrate that path-based solutions to the SUE problem with explicit path generation produce better results than traditional deterministic and stochastic algorithms with implicit path enumeration. Bekhor and Toledo (2005) discuss that, even though the correctness of link-based algorithms necessitates implicit choice sets, generated paths may be unrealistic from a behavioral perspective. For example, efficient paths that bring the driver closer to the destination and farther from the origin (e.g., Dial, 1971; Maher, 1998; Dial, 2006) may be unrealistic when involving repeatedly getting on and off a freeway, and cyclic paths from solutions to the traffic assignment problem (e.g., Bell, 1995; Akamatsu, 1996) are unlikely to be chosen by any traveler. Bliemer and Taale (2006) illustrate that a separate choice set generation step prior to the assignment calculations offers a number of theoretical and computational advantages for the traffic assignment part of travel demand prediction in large networks.

The purpose for path generation is the second aspect of the choice set generation problem considered in this review. Three major purposes may be individuated: (i) analysis of travel alternatives to determine their availability, number, characteristics, variety and composition; (ii) estimation of disaggregate demand models to uncover behavioral parameters of utility functions at the individual level, using observations of individual route choices; (iii) prediction of choice probabilities to determine route and link flow levels in networks, using route choice models with estimated parameters. Given these three purposes, the requirements of the generated choice sets are the third aspect of the choice set generation problem considered in this review.

From the perspective of the analysis of travel alternatives, statistics about route characteristics and measures of similarity distribution provide a picture of the alternatives available to travelers. Unfortunately, considered choice sets are rarely observed (e.g., Hoogendoorn-Lanser, 2005; Prato, 2005), even though these rare observations provide some insights. Travelers do not consider all feasible alternatives, but most likely limit the availability to attractive routes on the basis of their constraints, preferences and experiences. For example, some routes may not be considered because of the presence of loops, while others may not be perceived as distinct alternatives because of high overlap. Some routes may not be satisfactory because of a combination of constraints, while others may not be feasible because of unrealistic repeated passages between roads of different functional hierarchy (e.g., repeatedly entering and exiting from a highway).

From the perspective of the demand model estimation, Bovy (2009) argues that the requirements on the quality of the choice sets are not strict, as not all relevant alternatives have to be included since satisfactory results in terms of utility parameter estimates are obtained also for small well-sampled choice sets. However, Bekhor and Prato (2006), Prato and Bekhor (2007) and Bliemer and Bovy (2008) illustrate that choice set size and composition affect model estimates and choice probabilities for several model specifications. Apparently, the argument proposed by Bovy (2009) is correct only when MNL or its modifications are estimated, not when taste heterogeneity, substitution patterns or heteroscedasticity are introduced within Generalized Extreme Value (GEV) or Mixed models.

From the perspective of the prediction of route and link flows, Bovy (2009) argues that the requirements on the quality of the choice sets are very strict, as all relevant routes have to be included within the generated choice set. Moreover, Bovy (2009) adds that the inclusion of unattractive routes in the choice set is expected neither to distort the demand predictions nor to seriously influence the computational efficiency. However, an objective definition of relevant route in real networks is currently missing, thus certainty about the correctness of route choice sets for prediction purposes appears related to the sensitivity and experience of the analyst, rather than to objective measures of choice set quality.

Given these premises, this review focuses on explicit choice set generation methods that produce a subset of a universal set of feasible paths for model estimation and flow prediction purposes. Note that not only the universal set is extremely large and complex, but also the generated choice set of feasible and attractive routes is most likely large, heterogeneous in terms of route composition, complex in terms of physical overlap among the routes, and dependent on the loading patterns affecting route attribute values.

The classification of choice set generation methods relies on the distinction between deterministic shortest path-based methods, stochastic shortest path-based techniques, constrained enumeration algorithms and probabilistic approaches. Their evaluation depends on the satisfaction of the mentioned choice set requirements and on the results of case studies presented in the literature.

## **2.2 Deterministic shortest path-based methods**

The largest group of path generation methods is based on repeated shortest path searches in the network, where the computation of optimal paths follows the modification of one or more input variables such as link impedances, route constraints and search criteria. Common characteristics of these methods are that solutions are deterministic, origin-destination pairs are processed sequentially and most of the procedures are heuristic (except for  $K$ -shortest path algorithms).

### **2.2.1 Shortest path algorithms**

The most straightforward approach to the choice set generation problem consists in the calculation of the best  $K$  paths according to some link additive generalized cost function. The behavioral assumption behind the search for  $K$  shortest paths is that travelers limit their choices among a certain number of minimum cost paths and avoid extremely costly alternatives. The number of generated paths defines the window of admissible costs between the shortest path and the  $K$ -th shortest path, as a large

number provides larger availability of choices and includes much longer alternatives than the actual shortest path.

Existing algorithms extend the approach adopted to determine a single shortest path (e.g., Bellman, 1958; Dijkstra, 1959). Literature presents several exact solutions to the  $K$ -shortest path problem. Some algorithms allow cycles in the paths (e.g., Hoffman and Pavley, 1959; Bellman and Kalaba, 1960; Dreyfus, 1969; Shier, 1976; Shier, 1979; Eppstein, 1998), while other algorithms consider only acyclic paths (e.g., Yen, 1971; Lawler, 1976; Katoh et al., 1982; Hadjiconstantinou and Christofides, 1999).

For both estimation and prediction purposes, the limitation of the  $K$ -shortest path technique is the possibility of generating over circuitous and extremely similar routes that are highly unattractive to travelers. A possible solution consists in the implementation of a  $K$ -shortest path algorithm that considers acyclic paths with a large number  $K$  of routes. Bekhor et al. (2006) use the  $K$ -shortest path algorithm with  $K$  equal to 15 and 40 routes, and compare the generated paths with observed routes collected among M.I.T. faculty and staff members. Remarkably, the increase in the percentage of actual routes reproduced is equal to 1% when 25 routes are added. Also, 30% of the observed routes are not replicated as the algorithm apparently produces irrelevant routes. The increment of the number  $K$  of routes seems not to solve the problem of high similarity, otherwise the percentage of reproduced routes would increase significantly with the considerable increment of  $K$ .

Alternative solutions are proposed in the literature to capture preferences of different travelers by increasing route heterogeneity. Lombard and Church (1993) introduce the concept of “gateway shortest path”, by constraining paths through nodes (“gateways”) dispersed in the network to allow the generation of spatially distributed routes. Kuby et al. (1997) propose an iterative method that selects a subset among  $k$ -shortest paths, by minimizing a similarity measure based on the shared links between routes. Hunt and Kornhauser (1997) present the concept of “essentially least cost path”, by operating on a cost threshold to construct paths that contain locally acceptable detours from the shortest path. Akgün et al. (2000) treats path enumeration as a “ $p$ -dispersion” problem to generate spatially dissimilar routes, by applying a heuristic that searches for an initial  $k$ -shortest paths solution and performs a local search based on minimizing similarity to improve this initial solution. Van der Zijpp and Fiorenzo-Catalano (2005) propose an alternative method to find the feasible shortest paths according to a wide class of behavioral constraints, while maintaining the same computational efficiency of the other algorithms.

These techniques generate attractive acyclic and heterogeneous paths. However, their implementation suggests two limitations for estimation and prediction purposes, respectively. From the modeling perspective, all travelers moving between the same origin-destination pair share the same generated choice set, but this proposition appears unrealistic since a sound behavioral assumption should assure each traveler to have personal constraints, preferences and experiences leading to individual choice sets. From the predicting perspective, the absence of objective measures of route attractiveness leaves only to the sensitivity and experience of the researcher the control about whether all the attractive routes are generated.

## 2.2.2 Labeling approach

The behavioral assumption underneath the labeling approach is that travelers have different objectives. Some drivers may wish to minimize travel time, while others may

feel uncomfortable with heavily congested roads or freeways. Some drivers may prefer to drive through familiar landmarks, while others may look for scenic routes in absence of time constraints. Each criterion may correspond to a different preferred route and each route may be labeled according to a different objective function for which the path is optimum.

Ben-Akiva et al. (1984) propose a labeling approach in which physical paths form the upper-level nest of a Nested Logit (NL) model structure. Since travelers choose among physical routes and not among labels, the estimation setting of the NL model is modified with respect to the traditional setting in which the choice of the lowest-level alternative is known. The estimation of the labeling model consists of the analytical calculation of the inclusive values and the maximum likelihood estimation of the non-linear in parameters NL model. Several specifications of model parameters consider path attributes and label-combination dummies in the physical path nest, and only label-specific constants at the labeled path level.

Ramming (2002) presents a simplified version of the approach and searches for the shortest path according to sixteen labels, from the more conventional rule of searching for the minimum travel time to the more unconventional rule of looking for the maximum travel time in safe neighborhoods. Prato and Bekhor (2006) consider four attributes and evaluate the created paths with respect to chosen routes gathered among faculty and staff members of Turin Polytechnic. Both case studies reveal that the labeling approach reproduces only partially actual choices of actual drivers, and suggest that the unsatisfactory coverage highly depends on the definition of the labels. The discretion of the analyst guides the label definition, and the correct guess for the objective function for each traveler appears a complex task requiring a priori knowledge about travelers' preferences for an efficient implementation.

### **2.2.3 Link elimination**

The link elimination approach is based on the repetitive search for the shortest path after removal of part or all the shortest path links from previous searches. Behaviorally, this approach guarantees dissimilarity among alternative paths that obviously do not share links in the measure defined by the link elimination rule. These rules are left to the discretion of the researcher, who could eliminate for example either all the links in the current shortest path, a link that drives the traveler further from the destination and closer to the origin, or a link that belongs to a minor road rather than to a major arterial.

Bellman and Kalaba (1960) illustrate a heuristic similar to the elimination approach. The "branching" procedure calculates the shortest path, identifies links with one node on this path and one node off this path, and selects a link that satisfies a pre-defined heuristic rule (e.g., first link starting from the origin, link with the maximum capacity). A new path is formed by computing the shortest path from the head of the selected link to the destination, or from the origin to the tail of the selected link. Additional paths may be generated by using other links. Azevedo et al. (1993) present the original link elimination approach where the elimination rule removes all the links from the shortest path before searching for the next optimal path. The shortcoming of this method is the network disconnection, as the removal of centroid connectors and major crossings does not guarantee the existence of additional paths between origin and destination of a trip.

Bekhor et al. (2006), Prato and Bekhor (2006), Frejinger and Bierlaire (2007) apply a variant of the original approach, by calculating the minimum travel time path

and removing one link from the shortest path at each of fifty, ten and fifty iterations respectively. Only Prato and Bekhor (2006) detail the applied elimination rule, which removes one link when the final node is farther from the destination and closer to the origin with respect to the initial node. When compared with respect to actual choices of actual drivers, these three studies report between 60% and 80% of replicated observed routes.

The implementation of link elimination appears problematic from a behavioral perspective. When the original formulation is applied, the network disconnection problem does not allow generating all attractive routes using alternative access to major crossings, even though drivers often use alternative paths to a specific major crossing. When the single link elimination is performed, the method produces extremely similar routes, as for each removal of a link there is always a short deviation using links close to the removed one. The implementation of the variant of the original approach appears problematic also from a computational perspective, as realistic paths may contain a large number of links and the number of possible combinations of links to be removed may discourage the elimination of single links.

#### **2.2.4 Link penalty**

The link penalty approach is also based on the repetitive search for the shortest path, but a penalty on the impedance of all links in the resulting shortest path is imposed instead of the link removal. Behaviorally, also this method discourages the selection of the same set of links and attempts to generate routes that are more dissimilar as the amount of penalization increases.

De la Barra et al. (1993) introduce the link penalty heuristic, which constructs the choice set by iteratively storing paths. The first route stored is the actual shortest path, and the additional stored paths are the shortest paths computed after impedances on links belonging to the previous path are increased according to a fixed percentage. The procedure ends when two stored routes coincide. Ruphail et al. (1995) define the Iterative Penalty Method (IPM), where the penalization may be applied not only to the current shortest path, but to all the shortest paths calculated, by either adding a fixed impedance amount or multiplying their impedances by a fixed term. Park and Rilett (1997) propose a variant of the original approach to avoid producing minor deviations at the start or the end of the route. Specifically, impedances on links are not increased within a certain distance from the origin or the destination of the trip. Consequently, generated paths are less similar and the relevance of the routes increases. Scott et al. (1997) enrich the approach with an optimization program for determining the penalty factor for shortest path links in order to generate a next shortest path that overlaps with the current one by no more than a given number of links.

Bekhor et al. (2006) define penalty factors proportional to the distance between origin and destination of the observed routes, so that higher increments occur for longer distances. Prato and Bekhor (2006) identify a unique penalty factor and iterate the procedure fifteen times. Both studies show that link penalty does not provide improvement in terms of generation of relevant routes, as the method reproduces slightly inferior percentages of observed routes with respect to the link elimination approach.

The advantages of link penalty are the prospect for essential links to remain in the network and the discouragement for already identified links to belong to shortest paths in further iterations. The limitations of this approach consist in the generation of high impedance paths that undermines the relevance of the routes and in the dependency on



the definition of the penalty factor. With low values the algorithm is not computationally efficient as the same path is identified repeatedly, while with high values very high impedance and unattractive paths are generated before shorter and more attractive ones.

### **2.3 Stochastic shortest path-based methods**

A smaller group of path generation methods is based on repeated shortest path searches in the network, where the computation of optimal paths follows the random extraction of link impedances and individual preferences from probability distributions. Common characteristics of these methods are that solutions are stochastic, origin-destination pairs are processed simultaneously and all the procedures are heuristic. Moreover, stochastic path generation is generally a case of importance sampling because the selection probability of a route depends on the properties of the route itself, such as length or travel time. Accordingly, route choice models require a sampling correction term accounting for unequal selection probabilities of alternative routes in order not to produce biased parameter estimates (Frejinger, 2007; Bovy et al., 2009; Frejinger et al., 2009).

#### **2.3.1 Simulation approach**

The behavioral assumption underneath the simulation approach is that travelers perceive path costs with error, which is represented by extracting generalized cost functions from probability distributions.

The idea of the simulation approach originates from the application of the Multinomial Probit (MNP) model to traffic assignment. Sheffi and Powell (1982) design a procedure that, at each iteration, implements a Monte Carlo technique to draw link travel times from the Probit distribution around the overall congested cost function. Then, the procedure performs All-Or-Nothing assignment and computes the final link flows as the average values of the flows from all the iterations. Cascetta (2001) argues that the simulation of the distribution should be performed only around the non congested part of the cost function. Fiorenzo-Catalano and Van der Zijpp (2001) implement a version of the Monte Carlo technique by gradually increasing the variance of the random components in the model in order to keep at a constant rate the frequency with which new paths are found. Ramming (2002) proposes the replication of the same procedure for path generation purposes, only at each iteration the shortest path is saved instead of the link flows.

Nielsen (2000) extracts random draws from a truncated normal in a study concerning the Copenhagen-Ringsted railway project. Ramming (2002) extracts 48 draws from a normal distribution, with mean and variance equal to the link travel time. Bierlaire and Frejinger (2005) extract 20 draws from a truncated normal distribution, with mean and variance equal to the link travel times recorded by GPS units during a traffic safety study in the Swedish city of Borlange. Prato and Bekhor (2006) extract 35 draws from a truncated normal distribution, with mean and variance equal to the link travel time, left truncation limit equal to the free-flow time and right truncation limit equal to the travel time calculated for a minimum speed assumed equal to 10 km/h.

The advantages of the method are related to the generation of a large number of attractive routes, provided the selection of appropriate probability distribution and number of draws. Granted proper correction for unequal selection probabilities

(Frejinger, 2007; Bovy et al., 2009; Frejinger et al., 2009), stochastically generated choice sets are suitable for model estimation and also for flow prediction, conditional on the definition of a large number of routes that most likely will contain all the attractive alternatives.

The limitations of the method are related to the discretion in the selection of probability distribution and number of draws. The normal distribution may represent the distribution of the perceived costs, but negative draws must be truncated as link impedances are non-negative. Unlike the normal, the truncated normal distribution is not additive in mean and variance. In the application of simulation in the SUE problem, this property does not guarantee independency from the link segmentation (Nielsen, 2000) and distributions such as log-normal and gamma are preferred because non-negative draws are guaranteed (Nielsen and Frederiksen, 2006). In the application of simulation in single random draw extractions, the distribution does not necessarily need to be reproductive and truncated distributions are admissible, even though Nielsen (2000) discusses that the truncation leads to biases toward certain routes and gamma distribution is preferred.

Ramming (2002) extracts a number of draws that contains the computational costs within the same boundaries of link elimination and link penalty approaches, as the computational expenditure of the simulation approach is obviously proportional to the number of draws. Note that the increment in the number of draws significantly increases the computational costs, but does not guarantee the generation of additional unique paths. Prato and Bekhor (2006) extract a number of draws that exhausts the ability of the method to generate unique paths, as the standard deviation limits the possibility of producing highly different routes. With low values of the variance the algorithm is not efficient because few unique paths are generated repeatedly, while with high values the approach is also not efficient because numerous unrealistic paths are produced.

### **2.3.2 Doubly stochastic generation function**

The behavioral assumption of this approach is that not only travelers perceive path costs with error, but also different travelers have different perceptions. Accordingly, the generation function has a random term for the generalized cost function and a random term for the traveler taste heterogeneity.

Nielsen (2000) introduces the idea of double stochastic function by designing a heuristic modification of the SUE problem with a first component that considers road users' perception of the traffic network at the link level, and a second component that considers differences within road users' utility functions. Bovy and Fiorenzo-Catalano (2007) propose to utilize a trip utility function as the basis for a doubly stochastic generation function. Relevant routes are created through optimal path search in the given network by stochastically varying network attributes and attribute preferences. Variation in link impedances reflects differences among travelers in the perception of link attributes and network knowledge. Variation in the parameter values reflects differences among travelers in the preferences for these attributes.

Bliemer and Taale (2006) apply the generation function on the Dutch national main road network before the calculation of traffic flow predictions with a dynamic assignment model. Bovy and Fiorenzo-Catalano (2007) calibrate the variances and apply the approach to the Rotterdam-Dordrecht corridor in the Netherlands. The calibration of the function parameters involves a set of variance values to be

determined by maximizing some measure of resemblance with observations of individual consideration sets or chosen routes.

The advantages of the method are the heterogeneity of the generated alternatives, the relevance of the routes given the high conformity with observed choices, and the computational efficiency in large networks. The shortcoming of the technique consists in the calibration of the probability function coefficients, as consideration sets are difficult to collect and the use of incorrect values could produce irrelevant and unrealistic routes. Also, this technique deals with the same issues of the simulation approach with respect to the selection of the probability distributions, with the additional problem of the distribution of the value-of-time. Note that when cost and time are simultaneously considered, a normal distribution of the cost coefficient implies an unacceptable distribution of value-of-time with undefined mean and variance.

## 2.4 Constrained enumeration methods

Constrained enumeration methods rely on the behavioral assumption that travelers choose routes according to behavioral rules other than the minimum cost path.

Prato and Bekhor (2006) propose a branch and bound algorithm where the branching rule reflects behavioral assumptions through the definition of thresholds. A directional threshold excludes from consideration links that take the driver significantly farther from the destination and closer to the origin of the trip. A temporal threshold rejects paths that travelers would consider unrealistic since their travel time is excessively high. A loop threshold discards routes that travelers would not consider because they include large detours. A similarity threshold removes highly overlapping paths that travelers would not consider as separate alternatives. A threshold constraint removes unrealistic paths containing maneuvers causing delay in terms of travel time and apprehension in drivers approaching junctions (Prato and Bekhor, 2006).

The algorithm constructs the connection tree between origin and destination of a trip by processing sequences of links according to the branching rule that accounts for the logical constraints formulated to increase heterogeneity of the generated paths. Each sequence of links connecting origin and destination and satisfying all the constraints enters the choice set as a feasible solution to the path generation problem (Prato and Bekhor, 2006). The algorithm elaborates previous transport related applications of constrained enumeration techniques. Friedrich et al. (2001) apply a branch and bound assignment procedure for transit networks using a timetable-based search algorithm. Hoogendoorn-Lanser (2005) adapts the same procedure for choice set generation in the multimodal network context. In these applications the technique exploits predefined route sections, a highly restrictive assumption for path generation in road networks.

Prato and Bekhor (2006) show that the algorithm produces heterogeneous and relevant routes, as almost all the observed choices of actual travelers are replicated. Bekhor and Prato (2009) apply the algorithm to the Boston case study and confirm that behavioral assumptions other than the shortest path selection are more effective to produce realistic paths. From a computational perspective, the speed of the algorithm depends exponentially on the depth of the tree and consequently on the number of links in the paths. The implementation of the algorithm would accordingly be limited to small networks (see e.g., Frejinger, 2007), even though empirical results prove the

opposite when different design parameters are applied to the large Boston network with respect to the small Turin network (Bekhor and Prato, 2009).

For estimation purposes, constrained enumeration produces an exhaustive choice set extremely valuable for utility parameter estimation. For prediction purposes, the main issue remains the difficulty to determine whether all the relevant routes are generated. However, the exhaustive and exact nature of the solution suggests that all attractive routes should be generated with some unattractive routes, without significant problems for flow predictions (see Bovy, 2009). The shortcoming of the method is related to the definition of the thresholds of the behavioral constraints. The absence of a sensitivity analysis reflects the lack of knowledge about the effects of the variation of the behavioral thresholds.

## 2.5 Probabilistic methods

Probabilistic methods attach a generation probability to each route. The adoption of the full probabilistic approach proposed by Manski (1977) is impossible in route choice. In fact, even for a small choice set of ten routes (small in dense urban networks), calculating the selection probabilities of each of the 1023 potential choice sets appears impractical and unmanageable.

Cascetta and Papola (2001) propose the Implicit Availability/Perception (IAP) model, where the probability of choice set membership of an alternative enters the utility function of the choice model. Intuitively, a low membership probability of an alternative leads to a decrease of its choice probability. Even though the membership probability is supposed to depend on attitudinal and perception variables, the only successful known estimation of the model shows that this probability actually depends on socio-economic variables and utility attributes. Ramming (2002) details how any attempt to estimate IAP Logit with variables related to the network knowledge does not produce satisfactory results.

Frejinger (2007) and Frejinger et al. (2009) calculate a probability for each link in the network. For a given origin-destination pair, the probability of a link depends on its distance from the shortest path according to a generalized cost function. Accordingly, all links on the actual shortest path have a link probability of one, and other links between zero and one. Starting from the origin, a repeated random walk procedure adds links successively from node to node with the link selection process at each node governed by the probabilities of the associated next links. At the destination, the route probability corresponds to the product of the associated link probabilities and is used to correct the unequal sampling probability when the resulting route choice set is used for model estimation. Frejinger (2007) and Frejinger et al. (2009) apply the random walk to a synthetic network to illustrate the advantages of the generation method and the related importance sampling estimation. Results illustrate that models accounting for the sampling correction are remarkably better than the ones that do not include the correction for the designed synthetic network. However, the positive results of this application appear related to the unusual synthetic network designed in order to avoid long and cyclic paths, rather than to the effectiveness of the method. In fact, routes generated with random walks may be very circuitous, contain loops and be extremely long since they do not reach the destination in a reasonable number of steps, thus not being suitable for estimation and prediction purposes.

### 3. Route choice models

#### 3.1 The problem

Subsets of alternative routes generated with the described path generation techniques are usually quite large for both model estimation and flow prediction, since all the relevant routes are possibly included and some irrelevant routes are probably created. Case studies in European and American urban networks show that up to 70 alternatives are considered widespread procedure to estimate discrete choice models, and up to 100 alternatives are regarded as common practice to perform traffic assignment. Intuitively, the number of alternatives in the choice set plays a role in the estimation of discrete choice models within the route choice context.

Recent studies provide insight into the choice set effect on model estimates and performances. Bekhor et al. (2006) examine the effect of sample size in path-based assignment by examining the well-known Winnipeg network and showing that larger choice sets improve convergence and objective function values. Prato and Bekhor (2007) illustrate a significant influence of size and composition of generated route sets on parameter estimates and model performances, even though reported results discuss aggregate performance statistics rather than estimation quality at the level of individual choice sets. Frejinger (2007) shows that the full choice set is necessary to calculate correlation measures and obtain optimal estimation results. Bliemer and Bovy (2008) address the problem of route choice prediction and show that the addition of irrelevant routes biases route choice probabilities and causes attractive routes to become less attractive, in contrast with the expected model robustness toward the addition of irrelevant routes (Bovy, 2009).

Accordingly, route choice models should exhibit robustness in utility parameter estimates with respect to choice set size. For estimation purposes, this model requirement would allow the definition of choice sets with a reasonable number of attractive alternatives in order to obtain reliable model estimates. For prediction purposes, this model requirement would not solve the issue of the necessity to generate all the relevant routes.

Dense urban networks with 70 or 100 alternatives show a high degree of similarity among alternative routes. For this reason, most of the literature focuses on the correlation between alternatives, which alters choice probabilities of overlapping routes. Prashker and Bekhor (2004) illustrate the simple overlap problem, which is the route choice counterpart of the famous red bus - blue bus problem, and a simple route switching problem to observe that several route choice models do not correctly capture the similarity between routes in every network configuration. Bliemer and Bovy (2008) consider a Monte Carlo simulation of route choices from a grid network to illustrate that routes with the same length are not equally preferred, as the more the routes overlap, the lower their probability in favor of more independent alternatives. Clearly, route choice models should be able to represent properly the correlation structure among alternative routes. Whether the similarity is perceived at the link level or at the road hierarchical level is not certain, and models able to integrate both these aspects are preferable. Note that the most common discrete choice models in the practice of travel behavior modeling, Multinomial Logit (MNL) and Nested Logit (NL), are not suitable to model route choice. MNL does not allow to account for similarity among alternatives, while NL assumes that each alternative belongs

exclusively to one nest while in real-size networks routes share links with hundreds of other paths.

The last aspect concerns the prediction of route and link flows. SUE formulations are available for Logit and GEV models, not for enhanced models that account for complex similarity structures and potentially for taste heterogeneity and heteroscedasticity. Route choice models should be able to be translated into an equivalent mathematical formulation of a traffic assignment problem for prediction purposes.

The classification of different route choice models relies on the distinction between different model structures. Their evaluation depends on the satisfaction of the aforementioned model requirements and on the results of case-studies about real-size networks presented in the literature.

### 3.2 Logit structures

MNL-modifications maintain the simple Logit structure and introduce a correction term within the deterministic part of the utility function to approximate the correlation among alternative routes.

#### 3.2.1 C-Logit

Cascetta et al. (1996) propose the first modification of the MNL model, in which a commonality factor measures the degree of similarity of each route with the other routes in the choice set  $C$ . The expression of the probability  $P_k$  of choosing route  $k$  within the choice set  $C$  reflects the simple Logit structure of the model:

$$P_k = \frac{\exp(V_k + \beta_{CF} \cdot CF_k)}{\sum_{l \in C} \exp(V_l + \beta_{CF} \cdot CF_l)} \quad (1)$$

where  $V_k$  and  $V_l$  are the utility functions of route  $k$  and  $l$ , respectively,  $CF_k$  and  $CF_l$  are the commonality factors, and  $\beta_{CF}$  is a parameter to be estimated. Literature presents different formulations of the commonality factor (Cascetta et al., 1996, Cascetta, 2001):

$$CF_k = \ln \sum_{l \in C} \left( \frac{L_{kl}}{\sqrt{L_k L_l}} \right)^{\gamma_{CF}} \quad (2)$$

$$CF_k = \ln \sum_{a \in \Gamma_k} \left( \frac{L_a}{L_k} \sum_{l \in C} \delta_{al} \right) \quad (3)$$

$$CF_k = \sum_{a \in \Gamma_k} \left( \frac{L_a}{L_k} \ln \sum_{l \in C} \delta_{al} \right) \quad (4)$$

$$CF_k = \ln \left[ 1 + \sum_{\substack{l \in C \\ k \neq l}} \left( \frac{L_{kl}}{\sqrt{L_k L_l}} \right) \left( \frac{L_k - L_{kl}}{L_l - L_{kl}} \right) \right] \quad (5)$$

where  $L_k$  and  $L_l$  are the length of routes  $k$  and  $l$ , respectively,  $L_a$  is the length of link  $a$ ,  $\Gamma_k$  is the set of links belonging to route  $k$ ,  $L_{kl}$  is the common length between routes  $k$  and  $l$ ,  $\delta_{al}$  is the link-path incidence dummy, equal to one if route  $l$  uses links  $a$  and zero otherwise, and  $\gamma_{CF}$  is a parameter to be estimated.

The commonality factors express different concepts of similarity: expression (2) depends exclusively on the common length between routes, equations (3) and (4) extend the concept to all links in the route and introduce weights on the link importance proportional to the ratio between link and route lengths, expression (5) introduces also the costs of the non shared links and implies that the ratio between the commonality factors of two routes should increase when the overlapping between the two routes increases. The commonality factors in equations (2) through (5) are always positive. Consequently, the estimated parameter  $\beta_{CF}$  should be negative to express the reduction of the utility of paths with common links with respect to other routes. Also, for unique paths the arguments of the logarithms are equal to one and the commonality factors are null.

The advantages of the C-Logit model consist in the desired robustness of utility parameter estimates with respect to the choice set size (Prato and Bekhor, 2007) and in the existence of an equivalent mathematical formulation for the SUE problem (Zhou and Chen, 2003). The major disadvantages of the C-Logit model are that the commonality factor captures only part of the similarity and selection rules for the formulation of the commonality factors are not suggested. Case studies provide evidence of counter-intuitive results in terms of the sign of the parameter  $\beta_{CF}$  when the commonality factor is computed according to expressions (3) and (4), while illustrate expected and satisfactory results when the commonality factor is computed according to expression (5) (Ramming, 2002; Prato, 2005; Prato and Bekhor, 2007). Further, likelihood values show that the Path-Size Logit model generally outperforms the C-Logit model (Ramming, 2002; Prato and Bekhor, 2006; Prato and Bekhor, 2007). Most likely this explains why present research about MNL-modifications evolves toward new formulations of the path size correction term rather than new expressions of the commonality factor.

### 3.2.2 Path Size Logit

Ben-Akiva and Bierlaire (1999) present the Path-Size Logit (PSL) model for an application of discrete choice theory for aggregate alternatives, already used in other transportation contexts such as destination choice. Also for the PSL model the expression of the probability of choosing route  $k$  within the alternative paths reflects the simple Logit structure:

$$P_k = \frac{\exp(V_k + \beta_{PS} \cdot \ln PS_k)}{\sum_{l \in C} \exp(V_l + \beta_{PS} \cdot \ln PS_l)} \quad (6)$$

where  $PS_k$  and  $PS_l$  are the path sizes of routes  $k$  and  $l$ , respectively, and  $\beta_{PS}$  is a parameter to be estimated. Even though C-Logit and PSL have similar functional forms, each model gives a different interpretation with respect to the correction term introduced within the utility function. The commonality factor reduces the utility of a path because of its similarity with respect to other routes, while the path size indicates

the fraction of the path that constitutes a “full” alternative. Accordingly, a unique path has a size equal to one and  $N$  duplicate paths share the size  $1/N$ .

Literature presents different formulations for the path size (Ben-Akiva and Bierlaire, 1999; Ramming, 2002):

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{l \in C} \delta_{al}} \quad (7)$$

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{l \in C} \left( \frac{L_k}{L_l} \right)^{\gamma_{PS}} \delta_{al}} \quad (8)$$

where  $\gamma_{PS}$  is a parameter to be estimated.

The original path size formulation (7) expresses the weight corresponding to the fraction of path impedance coming from a specific link as the ratio between link and route lengths (Ben-Akiva and Bierlaire, 1999). Also, the remaining term is based on the number of paths using a specific link and therefore is equal to one for links used by only one path. Note that the impedance of paths using a specific link does not affect this term, and consequently the formulation can account for different size contributions due to routes with different lengths.

The generalized path size formulation (8) intends to decrease the influence of excessively long paths on the utility of shorter paths in the choice set (Ramming, 2002). Further, high values of the parameter  $\gamma_{PS}$  appear to increase the goodness-of-fit of the PSL model, as demonstrated by several case studies (Ramming, 2002; Prato, 2005; Hoogendoorn-Lanser, 2005; Bekhor and Prato, 2006). Frejinger and Bierlaire (2007) discuss that the generalized formulation may produce counter-intuitive results for high values of the parameter  $\gamma_{PS}$  and the original formulation is preferred, also because supported by a theoretical foundation.

The estimation of the PSL model with the original path size formulation is a simple task, given the closed-form Logit structure and the limited computational effort required for path size calculation. The estimation of the PSL model with the generalized path size formulation highly increases the computational costs, as the estimation of the parameter  $\gamma_{PS}$  requires the definition of a non-linear utility function. Only one study presents the actual estimation of the parameter  $\gamma_{PS}$  (Bekhor and Prato, 2006), while other studies obviate the computational problem by repeating the estimation of the PSL model for different values of the parameter  $\gamma_{PS}$  and individuating the optimal value (Ramming, 2002; Prato, 2005; Hoogendoorn-Lanser, 2005). Regardless of the computational issues, the behavioral interpretation of the estimates of the parameter  $\gamma_{PS}$  is extremely difficult given that the estimated or the optimal values vary between 10 and 15.

The limitation of the PSL model is that the path size captures only part of the correlation and makes preferable models that account for the correlation within the error structure without considerably increasing the complexity. Frejinger (2007) estimates the PSL model with sampling correction while considering the full choice set of paths as the actual choice set, and results show that unbiased estimates are obtained only when the correction term is calculated on the full choice set.



### 3.2.3 Path Size Correction Logit

Bovy et al. (2008) revisit the path size formulation by proposing the analytical derivation of a correction factor from random utility theory based on aggregate alternatives and from an approximation of GEV models such as Paired Combinatorial Logit and Cross Nested Logit. The resulting model is named Path Size Correction Logit (PSCL), as the correction factor replaces the original path size expression. Also for the PSCL model the expression of the probability of choosing route  $k$  within the alternative paths maintains the simple Logit structure:

$$P_k = \frac{\exp(V_k + \beta_{PSC} \cdot PSC_k)}{\sum_{l \in C} \exp(V_l + \beta_{PSC} \cdot PSC_l)} \quad (9)$$

where  $PSC_k$  and  $PSC_l$  are the path size corrections of routes  $k$  and  $l$ , respectively, and  $\beta_{PSC}$  is a parameter to be estimated. The derived formulation of the path size correction factor assumes the following expression (Bovy et al., 2008):

$$PSC_k = - \sum_{a \in \Gamma_k} \left( \frac{L_a}{L_k} \ln \sum_{l \in C} \delta_{al} \right) \quad (10)$$

The derived path size correction (10) weighs the length of the common links by the logarithm of the number of routes using these common links. The path size correction has an upper bound for completely independent routes, exactly as the other path size formulations, but does not have a lower bound because there is no upper bound on the number of paths sharing a link (Bovy et al., 2008). Note that the path size correction factor (10) varies between  $-\infty$  and  $0$ , while the original path size expression (7) varies between  $0$  and  $1$ .

The estimation of the PSCL is as simple as for the PSL, since the model maintains the closed-form Logit structure and the effort required for path size correction computation is limited. Model estimation for the Turin network and choice probabilities for a grid network show similar results to the original path size formulation, but the actual contribution of the formulation consists in its detailed and systematic derivation.

## 3.3 GEV structures

GEV models account for similarities within the stochastic part of the utility function and relate the network topology to the specific coefficients that characterize their tree structure, but do not allow to consider taste variation or correlation over time of unobserved factors.

### 3.3.1 Paired Combinatorial Logit

Prashker and Bekhor (1998) adapt to the route choice context the Paired Combinatorial Logit (PCL) model proposed by Chu (1989) and further developed by Koppelman and Wen (1998). Moreover, Prashker and Bekhor (2000) present a

mathematical formulation for the SUE problem for which the solution obtained is the PCL model.

The logic beneath the model is that routes are chosen among a pair of alternatives within the choice set, and the choice probability is defined accordingly:

$$P_k = \sum_{k \neq j} P(kl)P(k|kl) \tag{11}$$

where  $P(kl)$  is the marginal probability of choosing the pair  $(k,l)$  among the  $n(n-1)/2$  possible pairs, and  $P(k|kl)$  is the conditional probability of choosing route  $k$  given the chosen binary pair  $(k,l)$ . Conditional and marginal probabilities depend on the similarity between routes within the chosen pair:

$$P(k|kl) = \frac{\exp\left(\frac{V_k}{1-\sigma_{kl}}\right)}{\exp\left(\frac{V_k}{1-\sigma_{kl}}\right) + \exp\left(\frac{V_l}{1-\sigma_{kl}}\right)} \tag{12}$$

$$P(kl) = \frac{(1-\sigma_{kl}) \left( \exp\left(\frac{V_k}{1-\sigma_{kl}}\right) + \exp\left(\frac{V_l}{1-\sigma_{kl}}\right) \right)^{1-\sigma_{kl}}}{\sum_{p=1}^{n-1} \sum_{q=p+1}^n (1-\sigma_{pq}) \left( \exp\left(\frac{V_p}{1-\sigma_{pq}}\right) + \exp\left(\frac{V_q}{1-\sigma_{pq}}\right) \right)^{1-\sigma_{pq}}} \tag{13}$$

where  $\sigma_{kl}$  is the similarity coefficient between routes  $k$  and  $l$ .

The PCL model introduces independent similarity relationships for each pair of alternatives. The pair comparisons in the PCL model form the upper level of the nesting structure and therefore the number of nests increases rapidly with the network size as theoretically the upper level includes all possible route pairs. Literature presents two different formulations of the similarity coefficient. Prashker and Bekhor (1998) define a parameterized similarity index with some analogy with respect to the C-Logit factor presented in expression (2):

$$\sigma_{kl} = \left( \frac{L_{kl}}{\sqrt{L_k L_l}} \right)^{\gamma_\sigma} \tag{14}$$

where  $\gamma_\sigma$  is a parameter to be estimated. Gliebe et al. (1999) introduce a different expression for the similarity index, also depending on the common length of the routes in the pair:

$$\sigma_{kl} = \frac{L_{kl}}{L_k + L_l - L_{kl}} \tag{15}$$

Both equations (14) and (15) limit the similarity coefficient between zero and one, condition necessary for the PCL model to be consistent with random utility maximization. A coefficient equal to one indicates that all the links of a path are equal to the links of another path, while a coefficient equal to zero indicates that two paths have no link in common. If the latter condition is repeated for each  $(k,l)$  pair, all the routes in the choice set are disjointed and the PCL model collapses to the MNL model.

Note that literature reports PCL predictions of route choices per population segments in a multi-modal context (Benjamins et al., 2002) and the evaluation of the influence of choice set composition on PCL choice probabilities (Bliemer and Bovy, 2008). However, literature does not report examples of PCL estimation for real size networks, as the large number of pairs of alternatives enlarges the number of nests and significantly increases the computational expenditure.

### 3.3.2 Cross Nested Logit

Prashker and Bekhor (1998) adapt to the route choice situation the Cross Nested Logit (CNL) model proposed by Vovsha (1997) for a mode choice application in a large metropolitan area. Further, Prashker and Bekhor (2000) give an equivalent mathematical formulation of the CNL model for the SUE problem.

The assumption underneath the model is that routes are chosen within nests, which physically correspond to the links in the network, and the choice probability is defined accordingly:

$$P_k = \sum_m P(m) P(k|m) \tag{16}$$

where  $P(m)$  is the marginal probability of choosing a nest  $m$ , and  $P(k|m)$  is the conditional probability of choosing route  $k$  in nest  $m$ . Conditional and marginal probabilities depend on inclusion and nesting coefficients:

$$P(k|m) = \frac{(\alpha_{mk} \exp(V_k))^{1/\mu_m}}{\sum_l (\alpha_{ml} \exp(V_l))^{1/\mu_m}} \tag{17}$$

$$P(m) = \frac{\left( \sum_k (\alpha_{mk} \exp(V_k))^{1/\mu_m} \right)^{\mu_m}}{\sum_h \left( \sum_k (\alpha_{hk} \exp(V_h))^{1/\mu_m} \right)^{\mu_m}} \tag{18}$$

where  $\alpha_{mk}$  are inclusion coefficients ( $0 \leq \alpha_{mk} \leq 1$ ) and  $\mu_m$  are nesting coefficients ( $0 \leq \mu_m \leq 1$ ). Inclusion coefficients represent the percentage of the generic link  $m$  used by the generic alternative route  $k$ , and are subject to a regularity constraint:

$$\sum_m \alpha_{mk} = 1 \tag{19}$$

Prashker and Bekhor (1998) define a functional relationship for the inclusion coefficient with respect to the links in a route:

$$\alpha_{mk} = \frac{L_m}{L_k} \delta_{mk} \quad (20)$$

where  $L_m$  is the length of link (nest)  $m$ ,  $L_k$  is the length of route  $k$ , and  $\delta_{mk}$  is the link-path incidence dummy, equal to one if route  $k$  uses link  $m$  and zero otherwise. Further, Prashker and Bekhor (1998) assume that all the links share a common nesting coefficient  $\mu$  to be estimated. Note that the CNL collapses to the MNL model when the nesting coefficient  $\mu$  is equal to one, while becomes probabilistic at the higher level and deterministic at the lower level when the nesting coefficient  $\mu$  approaches zero.

Examples of CNL estimation exhibit some computational and behavioral issues. From a computational perspective, commensurate with its complexity CNL requires longer estimation time than other model specifications (Ramming, 2002; Prato, 2005). From a behavioral perspective, the estimated nesting coefficient often approaches one and suggests that the CNL model tends to collapse to MNL (Ramming, 2002; Prato, 2005; Prato and Bekhor, 2006). Note that in these examples CNL does not outperform MNL-modifications, and the nesting coefficient approaching one implies that the similarity among the alternatives is not properly captured by the definition of the inclusion coefficients. Accordingly, the model performance does not meet the expectations.

### 3.3.3 Generalized Nested Logit

Bekhor and Prashker (2001) adapt to the route choice framework the Generalized-Nested Logit (GNL) model, elaborated by Wen and Koppelman (2001) as a generalization of the CNL specification presented by Vovsha (1997). The same logic of the CNL model applies to the GNL model, and the choice probability is equally formulated according to expressions (16), (17) and (18). Moreover, Bekhor and Prashker (2001) also write the equivalent mathematical formulation of the model for the SUE problem.

Bekhor and Prashker (2001) generalize the CNL model by maintaining the expression (20) for the inclusion coefficient and allowing each nest  $m$  to have a different nesting coefficient. The proposed formulation assumes that the nesting coefficient is a parameterized average value of the inclusion coefficients:

$$\mu_m = \left( 1 - \frac{\sum_{l \in C_n} \alpha_{ml}}{\sum_{l \in C_n} \delta_{ml}} \right)^\gamma \quad (21)$$

where  $\gamma$  is a parameter to be estimated.

Estimation of GNL models shows similarities with respect to CNL models. The additional complexity from the estimation of the parameter  $\gamma$  further enlarges the computational effort required. Moreover, the estimated values of this parameter are usually relatively high, between 4 and 5 (Ramming, 2002; Prato, 2005; Prato and Bekhor, 2006): note that the nesting coefficient tends to one as the parameter

increases, and consequently the GNL exhibits the tendency to collapse to the MNL model exactly as the CNL in real case studies.

### 3.4 Non-GEV structures

Non-GEV model structures allow not only unrestricted substitution patterns, but also random taste variation and correlation in unobserved factors over time. Since these models do not present closed-form expression for the choice probabilities, Maximum Simulated Likelihood is required for their estimation.

#### 3.4.1 Multinomial Probit

Daganzo and Sheffi (1977) propose the MNP route choice model by assuming a normal distribution for the random component. The joint density function of the error terms is described by a (*J-length*) vector of means and a ( $J \times J$ ) covariance matrix, where *J* is the number of routes for a specific origin-destination pair.

The cumulative normal distribution function cannot be expressed in closed form, therefore the calculation of the probit choice probability of choosing a route among a large number of alternatives is not straightforward. The main problem of the MNP model is related to the specification of the covariance matrix, necessary to calculate choice probabilities, which needs to relate the variance of the error terms to measurable network parameters.

Sheffi and Powell (1982) assume that the variance is proportional to a fixed characteristic of the link (e.g., length), in order to obtain an invariant distribution of the error terms. Yai et al. (1997) propose a MNP model with a structured covariance matrix to represent the overlapping between route alternatives. The covariance matrix is related to measurable overlapping variables, such as the common length among the routes, and the stochastic loading is performed by numerical integration. Note that this assumption is essentially the same of Sheffi and Powell (1982), only within a different context. In fact, Yai et al. (1997) estimate route choice models using structured covariance matrices with fixed costs, while Sheffi and Powell develop the variance matrix in the context of their SUE program to show its equivalence to the SUE problem (Sheffi, 1985).

The costly computational effort for estimating MNP model explains why route choice modeling and stochastic assignment often adopt alternative specifications for representing travelers' behavior.

#### 3.4.2 Logit Kernel with Random Coefficients

The defining characteristic of the Logit Kernel (LK) or Mixed Logit model (Ben-Akiva and Bolduc, 1996; McFadden and Train, 2000) is that the unobserved factors can be decomposed into a part that contains correlation and heteroscedasticity, and another part that is i.i.d. extreme value.

The most straightforward derivation of the LK model is based on random coefficients. The probability for an individual *n* of choosing route *k* has the same form of the standard Logit, but it is conditional on the distribution of the coefficients  $\beta_n$ :

$$P_{nk}(\beta_n) = \frac{\exp(\beta'_n X_{nk})}{\sum_{l \in C_d} \exp(\beta'_n X_{nl})} \quad (22)$$

where  $X_{nk}$  and  $X_{nl}$  are observed variables that relate to the individual  $n$  and the alternatives  $k$  and  $l$ , respectively, and  $\beta_n$  is a vector of random coefficients representing the tastes of individual  $n$ . Since the modeler does not know  $\beta_n$  and cannot condition on  $\beta_n$ , the unconditional probability of choosing route  $k$  is the integral of  $P_{nk}(\beta_n)$  over all possible values of  $\beta_n$ :

$$P_{nk} = \int \frac{\exp(\beta' X_{nk})}{\sum_{l \in C_d} \exp(\beta' X_{nl})} f(\beta) d\beta \quad (23)$$

where  $f(\beta)$  is the density of the distribution of  $\beta$  over the population. The unconditional probability is computed by simulation:

$$P_{nk} = \frac{1}{D} \sum_{d=1}^D \frac{\exp(\beta'_d X_{nk})}{\sum_{l \in C_d} \exp(\beta'_d X_{nl})} \quad (24)$$

where  $\beta_d$  indicates a draw  $d$  from the distribution of  $\beta$  and  $D$  is the number of draws.

Despite the ongoing discussion about the suitability of coefficient distributions by researchers interested in modeling route choice behavior and solving traffic assignment, only a few studies actually illustrate estimates of route choice models accounting for taste heterogeneity and heteroscedasticity. Ben-Akiva et al. (1993) estimate a model with two alternatives and cross-sectional data where the time coefficient is log-normally distributed and the integral is evaluated with a Gaussian quadrature. Dial (1997) considers within a traffic assignment algorithm that drivers have different perceptions about travel times and costs because of habit, taste or information. Nielsen (2000) discusses that log-normal and gamma distributions are suitable to simulate preferences while proposing a stochastic traffic assignment model with differences in passengers' utility functions. Han et al. (2001) use uniform and normal distribution for delay and travel time in high traffic conditions to model SP games of pair wise route choices, and add that the log-normal distribution does not produce satisfactory results. Jou (2001) investigates the impact of pre-trip information on route choice behavior by estimating a random component model with normally distributed travel time. Lam and Small (2001) model the choice between a free and a tolled route by accounting for median travel time and variability of travel time according to the time of day. Nielsen et al. (2002) estimate a model for different driver categories by testing normally and log-normally distributed coefficients for travel time and cost. Nielsen (2004) evaluates heterogeneity in drivers' responses to pricing schemes by proposing a SP experiment about road pricing and modeling revealed preferences collected through GPS devices.

Given that travel time and cost coefficients are expected to be negative, the log-normal distribution should be preferred to the unbounded normal distribution that implies some travelers may actually prefer longer and costly trips. Results from case

studies illustrate that models with log-normal coefficients either do not produce satisfactory results (Han et al., 2001) or introduce excessive variation and sometimes produce illogical choices (Nielsen et al., 2002), while gamma distributed coefficients are preferable because reproductive and non-negative (Nielsen, 2000). Different discrete choice applications introduce bounded distributions for coefficients of variables with an expected sign (see, e.g., Train and Sonnier, 2005; Hess et al., 2005), but literature in the route choice context does not report the investigation of these distributions.

In the route choice context, models with varying cost coefficients are actually different from models with varying time coefficients, as the overall variance changes with the cost in the former case and with the time in the latter, causing the assumptions about heteroscedasticity to be different (Nielsen et al., 2002). Caussade et al. (2005) model heteroscedasticity in route choice while examining the influence of stated choice design complexity on consumer's ability to choose. The investigation of heteroscedasticity in the route choice context misses actual dedicated modeling of revealed choice behavior.

### 3.4.3 Logit Kernel with Factor Analytic Approach

A LK model can be used without a random-coefficients interpretation, as simply representing error components that create correlations among the utilities for different alternatives.

The first adaptation to route choice of the LK model assumes that the covariance of path utilities is proportional to the length by which paths overlap (Bekhor et al., 2002). Building on the derivation of the LK model with factor analytic approach, the probability of choosing route  $k$  given a vector  $\zeta$  of standard normal variables is formulated as follows:

$$P_k = \Lambda(k|\zeta) = \frac{\exp(\mu(X_k\beta + F_kT\zeta))}{\sum_{l \in C_n} \exp(\mu(X_l\beta + F_lT\zeta))} \quad (25)$$

where  $\beta_{(1 \times B)}$  is the column vector of parameters,  $X_k$  is the  $k$ -th row of the matrix of explanatory variables  $X_{(J \times B)}$ ,  $F_k$  is the  $k$ -th row of the factor loadings matrix  $F_{(J \times M)}$  ( $J$  paths and  $M$  network elements),  $T_{(M \times M)}$  is a diagonal matrix of covariance parameters  $\sigma_m$ ,  $\zeta_{(M \times 1)}$  is a vector of standard normal variables. Bekhor et al. (2002) assume that the link-specific factors are i.i.d. normal, the variance is proportional to the link length, the  $F$  matrix corresponds to the link-path incidence matrix and the  $T$  matrix corresponds to the link-factor variance matrix. Accordingly, the covariance parameter  $\sigma$  shared by each link is estimated.

The second adaptation to route choice of the LK model with error components proposes a different and interesting perspective (Frejinger and Bierlaire, 2007). The model assumes that two paths going through the same sub-network component may share unobserved attributes and be similar, even without sharing any link. Frejinger and Bierlaire (2007) define the  $F$  matrix as the sub-network component-path incidence matrix and the  $T$  matrix as the covariance matrix associated to the sub-network components. Accordingly, unique covariance parameters  $\sigma_m$  for each sub-network component are estimated.

LK models are demanding in terms of computational time, as the vector  $\zeta$  is unknown and the unconditional probability is computed by simulation:

$$P_k = \int_{\zeta} \Lambda(k|\zeta) \prod_{m=1}^{M_n} \varphi(\zeta_m) d\zeta_m = \frac{1}{D} \sum_{d=1}^D \Lambda(k|\zeta^d) \quad (26)$$

where  $\Phi(\zeta)$  is the standard normal density function,  $\zeta_d$  indicates a draw  $d$  from the distribution of  $\zeta$ , and  $D$  is the number of draws.

Other issues concern both adaptations of the LK model with factor analytic approach. When the covariance parameter shared by each link in the choice set is estimated, Ramming (2002) demonstrates instability of the estimates even with large number of draws and Prato (2005) discusses the difficulty of obtaining significant estimates. When the covariance parameters of each sub-network component are estimated, covariance estimates and model performances appear highly dependent on the discretion of the analyst in the selection of these components. Systematic rules for the definition of the sub-network components are not provided and the estimation of test models is required to individuate the most significant and appropriate structure of the  $F$  and  $T$  matrices. Accordingly, the computational cost increases significantly because of the test model estimation, even with low number  $D$  of random draws for the simulation of the integral.

## 4. Future research directions

This review provides lessons from the past about behavioral assumptions and motivations for developing choice set generation methods and route choice models, and also suggests directions for the future. As the description of the disadvantages of each method and model and the failure to satisfy the requirements suggest, several questions are undisclosed in modeling route choice behavior. This section focuses on three major areas of intervention: data collection of actual behavior, choice set generation of relevant alternatives, and discrete choice modeling of both choice set formation and choice process.

### 4.1 Data collection

Difficulties in data collection are probably the main reason behind the limited number of revealed preferences studies in the route choice context. Not only obtaining information about actual route choice behavior is challenging, but also transforming observed routes in link-by-link representations on a network is demanding.

Route choice data may be collected actively through interviews. Ben-Akiva et al. (1984) use data collected by stopping cars at the road side and handing or mailing questionnaires to the owner of the car, even though information focuses on “labels” rather than actual choice behavior. Ramming (2002) analyzes information about actual route choice behavior gathered by asking travelers to describe a chosen path through a sequence of route segments. Missing or dubious information through shortest path calculation between known path segments. Prato (2005) examines data about actual route choices of drivers by demanding them to fill a web-based form after recognizing on a web-based map their home-to-work chosen path. The existence of incorrect reporting of the network nodes is acknowledged to cause loss of valuable information. Vrtic et al. (2006) describe information about long-distance trips by requesting



travelers to describe the names of the origin and destination cities as well as maximum three intermediate cities or locations that they passed through. The difficulties of the modeler in correcting missing information, deciding to complete or eliminate partial path descriptions and matching the collected data with the network representation are issues that discourage the use of interviews.

Route choice data may be collected passively through GPS devices or cellular phone technologies. Several route choice studies use GPS data (e.g., Murakami and Wagner, 1999; Jan et al., 2000; Nielsen, 2004; Wolf et al., 2004; Li et al., 2005), while at the moment cellular phone technologies are only used for traffic detection and speed profile analysis (e.g., Asakura and Hato, 2004; Bar-Gera, 2007). Passive approaches present issues about data manipulation in order to obtain network paths. Data accuracy is limited by technical issues, likely related to the satellite coverage and to the receiver characteristics, as well as constrained by environmental issues, possibly related to the presence of obstacles to the reception or to the atmospheric conditions. Nielsen (2004) observes that an astounding 90% of the trips collected in the Copenhagen region are inaccurate because of missing information.

The inaccuracy for reported and GPS recorded trips potentially introduces biases and errors in model estimates. Bierlaire and Frejinger (2008) propose a solution to the problem by estimating route choice models within a framework that reconciles network-free data with a network based model. As an observed route is a sequence of individual pieces of data related to an itinerary, such as a sequence of GPS points, the Domain of Data Relevance (DDR) is defined as the physical area where each piece of data is relevant (Bierlaire and Frejinger, 2008). The definition of the DDR area depends on the associated concept and its extension is inversely proportional to the fuzziness of the concept (e.g., “downtown” is a larger area than “intersection between Connecticut Avenue and T street NW in Washington, D.C.”). The case study elaborates information about trips that contain details about origin, destination and a maximum of three intermediate locations (Vrtic et al., 2006), and the definition of the DDR of each location corresponds to the area zip code. When linking the network-free data with the network through the DDRs, the precision level of the observations corresponds to the precision level of the network, consequently the transportation network is simplified, especially in urban areas (Bierlaire and Frejinger, 2008). The estimation of a sub-network components model (Frejinger and Bierlaire, 2007) illustrates more accurate estimates after less time consuming data manipulation.

The network-free data approach could open possibilities for future research in the data collection area. The design of interviews could be simplified, not only for application in the interurban context as seen in the aforementioned study, but also in the urban context where trip description would be possible by simply collecting origin, destination and significant locations that psychologists and geographers define as landmarks (see, e.g., Gale et al., 1990; Friendschuh, 1992; Garling and Golledge, 2000). The adoption of passive collection techniques could also be simplified, even though advancements in GPS technology and cellular phone coverage are still required to limit data inaccuracy in order to correctly identify significant landmarks that would correspond to the DDRs.

Missing path segments, coding errors, approximation in the individuation of a specific point would not influence the data processing phase because of the simplifications proposed by the network-free approach, which would transform traditional or web-based interviews and passive collection methods into valuable sources of information. Also, the significant cost reduction of this phase would allow to collect a large amount of valuable information about actual route choices, and also

about aspects not fully explored such as their variability across different hours during the day, different days during the week, different trip motives and different population groups. Last, the eventual availability of actual route choices from the same drivers across a certain amount of time would provide an important comparison term for evaluating choice set generation methods not only with respect to single choices, but also with respect to observed sets of considered alternatives.

## 4.2 Choice set generation

The identification of relevant routes is the main reason behind the effort toward the development of choice set generation methods that are operational in large networks.

Modifications of the shortest path algorithms attempt to solve the problem of the excessive similarity among k-shortest paths by generating heterogeneous paths according to different measures of dissimilarity (Lombard and Church, 1993; Kuby et al., 1997; Hunt and Kornhauser, 1997; Akgün et al., 2000). Constrained enumeration methods introduce behavioral thresholds to express constraints within a k-shortest path algorithm (Van der Zijpp and Fiorenzo-Catalano, 2005) and a branch and bound technique (Prato and Bekhor, 2006). Heuristic techniques avoid the repetition of the same link sequences within the iterative search for the shortest path (Azevedo et al., 1993; De la Barra et al., 1993). The selection of similarity measures, the definition of behavioral thresholds and the designation of link elimination or penalization rules heavily depend on subjective rather than objective elements, given the importance of network and population knowledge of the researcher.

Stochastic shortest path-based methods attempt to generate heterogeneous routes by considering sources of uncertainty in the route choice context. Some techniques consider variability in the perception of route characteristics (Ramming, 2002; Prato, 2005; Bekhor et al., 2006), while other methods also consider variability in the preferences of travelers (Nielsen, 2000; Bovy and Fiorenzo-Catalano, 2007). The selection of the most suitable probability distribution and the assumption about its parameter values heavily depend on subjective considerations, and once more on the network and population knowledge of the researcher.

The aforementioned choice set generation methods consider the choice set formation either as a non-compensatory process on the basis of behavioral constraints or a preference driven process on the basis of the travelers' most important choice factors. Future research in choice set generation could consider that most likely choice set formation is a learning process through dynamic adaptations of cognition and perception of available and feasible options (Richardson, 1982). The availability of information regarding the variability of route choices across time for the same travelers, combined with dedicated surveys about driving behavior and attitudes, would allow to provide insight into this learning process.

Future research in choice set generation could also consider that choice sets are unobservable. Given also that deterministic constraints highly depend on the discretion and experience of the researcher, future research could move toward the definition of probabilistic constraints. Stochastic choice set generation models could capture the latent nature of the considered choice set, and could take inspiration from random constraint approaches (see Swait and Ben-Akiva, 1987; Ben-Akiva and Boccara, 1995; Swait, 2001; Basar and Bhat, 2004). Bovy (2009) illustrates that random constraint approaches are applicable for both estimation and prediction purposes, provided that the constraints are transformed from an individual to a more general level in prediction applications. Bovy (2009) also argues that random

constraint approaches are limited to simple choice situations with only a few alternatives, and that the adoption of a full probabilistic approach is nearly impossible in the route choice context because of the high number of possible alternatives and the consequent excessively high number of possible choice sets. Even though these considerations are correct, a stochastic choice set generation model appears the most natural evolution in this area of route choice behavior analysis. A non-compensatory approach prior to the stochastic model would limit the computational burden by eliminating highly unreasonable and irrelevant alternatives.

With respect to the evaluation of choice set generation methods, further research could focus on the related definitions of relevant routes and efficient choice set generation methods. Actual measures either indicate the percentage of observations for which an algorithm either reproduces observed routes according to a certain overlap threshold (Ramming, 2002) or exactly replicates the observed routes (Bovy and Fiorenzo-Catalano, 2007). These measures determine the level of plausibility of the generated routes because, if the observed choice is reproduced, the path set appears to be consistent with the observed behavior. However, these measures are not indicative of the efficiency of the algorithm, as they are not influenced by the number of alternatives generated. For example, a method that requires 50 iterations to cover 80% of the observations appears as good as a method that requires 10 iterations to cover the same amount of observations. Clearly, the latter is more efficient from the computational perspective.

The general consensus among route choice modelers indicates in the generation of relevant routes the objective to pursue for avoiding biases and errors in model estimates. However, relevant routes are easy to identify in synthetic network experiments (e.g., Bliemer and Bovy, 2008), but may be difficult to recognize in actual large networks. Relevance measures would answer the question whether a generated route is relevant, and efficiency measures would account not only for the ability to replicate actual route choice behavior, but also for the capacity to generate the most relevant set of alternative routes. Further, as anticipated in the data collection section, relevance and efficiency measures would largely benefit from the collection of actual route choices from the same drivers throughout a certain amount of time, which would provide invaluable insight into the routes actually considered by travelers and their switching patterns for the same origin-destination pairs.

### **4.3 Discrete choice models**

The representation of the correlation structure among alternative routes that share network links is the main reason behind the effort toward the adaptation of discrete choice models to the route choice context. Taste heterogeneity and heteroscedasticity are also analyzed in the literature, but at a lesser extent.

Approximation of the correlation structure within the deterministic part of the utility function is largely discussed in the literature. Researchers propose the derivation of correction terms from random utility theory based on aggregate alternatives and on the approximation of higher order GEV models (Frejinger and Bierlaire, 2007; Bovy et al., 2008). The common assumption is that path utilities decrease with increasing similarity with other routes in the choice set (Cascetta et al., 1996; Ben-Akiva and Bierlaire, 1999; Cascetta, 2001). Hoogendoorn-Lanser et al. (2005) discuss that the correlation does not lead to utility reduction when the degree of correlation among the alternatives in the choice set is comparable. Frejinger (2007) calculates the correction factor for the full choice set of possible alternatives to

demonstrate that the current procedure of computing the same factors for the generated choice set introduces biases in the model estimates. The approximation of the correlation structure is dependent on the specification of the choice set, and future research should address the issue of which approximation of the full choice set is suitable for the calculation of the correction factor in large networks.

Approximation of the correlation structure within the stochastic part of the utility function of GEV models is less discussed in the literature (Prashker and Bekhor, 1998; Bekhor and Prashker, 2001). Nesting coefficients for CNL and GNL models sometimes tend to one and these models tend to collapse to the MNL model and to perform worse than MNL-modifications (see Ramming, 2002; Prato, 2005; Prato and Bekhor, 2006). A nesting coefficient significantly different from one is obtained for a CNL model with a choice set generated with the branch and bound algorithm (Prato and Bekhor, 2007). These results apparently confirm the correctness of the assumption proposed by Hoogendoorn-Lanser et al. (2005), as the constrained enumeration method generates more heterogeneous routes, the correlation among alternatives is different and the CNL outperforms the MNL. Nonetheless, further research is required into the definition of the relationship between model parameters and network topology in GEV models. Specifically, the relaxation of the assumption that all the links are equally important would simplify the model structure and would encourage a realistic representation of route choice behavior, similarly to the idea proposed by Frejinger and Bierlaire (2007).

Future research is required to accommodate both substitution patterns and taste heterogeneity, which are usually considered separately. Some studies focus on the taste variation across the population by assuming preferably normal or log-normal distribution of time and cost (Ben-Akiva et al., 1993; Nielsen, 2000; Han et al., 2001; Jou, 2001; Nielsen et al., 2002; Nielsen, 2004), while other studies concentrate on the representation of the correlation structure (Bekhor et al., 2002; Frejinger and Bierlaire, 2007). Hess et al. (2006) discuss that this separation is reasonable at the conceptual level, but the risk of confounding is significant in practice, especially given the lack of knowledge about the true nature of the error structure. A combined error-components and random-parameters formulation would overcome the problem, provided that researchers make an informed choice of specification by performing a designed test (Hess et al., 2006). Alternative solutions to avoid the confounding problem could consist in the estimation of either a MNL-modification with random parameters or a GEV mixture structure. Random parameters in the utility function would account for taste variations, while the correction terms in the first model specification and the nesting structure in the second would represent substitution patterns. These alternative approaches would not be exempted by the mentioned problems related to the approximation of the deterministic correction of the utility function and to the expression of the nesting structure as a function of the network topology.

Future research should also address random constraint approaches to represent the choice process. Cascetta and Papola (2001) propose the concept of implicit availability perception within a Logit model structure, by considering that travelers could be unaware of some alternatives and modeling the availability of alternatives as a function of travelers' characteristic. Cantillo and Ortuzar (2005) and Cantillo et al. (2006) formulate a discrete choice model that incorporates thresholds in the perception of changes in attribute values, by allowing for random thresholds to differ between individuals and to be a function of socio-economic characteristics and choice conditions. Frejinger (2007) and Frejinger et al. (2009) model route choice with a sampling approach that considers the probability of selecting each path within the

universal choice set. Random constraint models could be able to account for the availability of alternative routes according to their attractiveness and also to latent personal traits that most likely influence route awareness (for example habit, network familiarity, and travel time perception). The main problem of these models is that the aforementioned formulations assume full knowledge about the universal choice set, which is possible to define in the mode choice context (Cascetta and Papola, 2001) and in synthetic experiments (Cantillo et al., 2006; Frejinger, 2007; Frejinger et al., 2009), but nearly impossible to identify in the route choice context. However, the problem of identification of the full set of alternatives would be overcome by the more feasible and behaviorally sound enumeration of the relevant routes, possibly defined according to route relevance measures.

## 5. Summary

This paper reviews path generation techniques and discrete choice models applied within the route choice modeling context. Even though the literature illustrates mainly the large effort in the development of enhanced probabilistic route choice models, the growing recognition of the importance of choice set size and composition motivates the equal importance given to path generation and model estimation in this retrospective. In addition to the description and the classification of existing methods and models for modeling route choice behavior, this paper considers their answer to requirements related to different purposes and their implementation within revealed preferences case studies.

Among choice set generation methods, constrained enumeration and doubly stochastic generation function demonstrate the required property of generating a large number of relevant routes by reproducing actual choices from actual travelers. Hence, both path generation methods illustrate suitability to both model estimation and flow prediction. Most likely, constrained enumeration techniques benefit from avoiding the assumption of the shortest path search lying beneath every other method. Probably, the doubly stochastic generation function benefits from accounting for every source of variability related to both route characteristics and travelers' preferences. Surely, the search for the shortest path is not the most effective behavioral assumption within the path generation framework.

Among route choice models, advantages and disadvantages may be individuated for each reviewed specification. From a computational perspective, MNL modifications are not challenging, G.E.V. models are more demanding because of the estimation of inclusion and nesting coefficients within complex model structures, and LK models introduce additional complexity because of the necessary simulation for choice probability calculation and the absence of an equivalent mathematical formulation of the SUE problem. From a behavioral perspective, MNL modifications depend on the definition of correction terms that may produce counter-intuitive results, while G.E.V. and LK models depend on theoretical formulations of the correlation structure among alternative routes. G.E.V. models seem preferable because of the superior theoretical foundation with respect to the MNL-modifications and the inferior computational complexity with respect to the LK models. Nonetheless, the definition of inclusion and nesting coefficients requires to be cautiously reconsidered as the assumption that all links are equally important to travelers appears unrealistic. Moreover, MNL-modifications seem preferable because of the stability of utility parameter estimates with respect to the variation of the choice set size.

This paper envisions future research directions in the three areas of data collection, choice set generation and discrete choice models. Future research in route choice modeling should also consider that not all route choice models are applied in stochastic assignment models. Equivalent mathematical formulation of the SUE problem are provided for MNL (Fisk, 1980), MNP (Sheffi and Powell, 1982), PCL (Prashker and Bekhor, 2000), CNL (Prashker and Bekhor, 2000), GNL (Bekhor and Prashker, 2001) and C-Logit (Zhou and Chen, 2003). The development of enhanced models should ultimately account for the possibility of writing an equivalent mathematical formulation of a traffic assignment problem.

## Acknowledgements

The insightful comments of three anonymous reviewers provided invaluable help in the revision of a previous version of this paper.

## References

- Akamatsu, T., 1996. Cyclic flows, Markov process and stochastic traffic assignment. *Transportation Research Part B*, 30(5), 369-386.
- Akgün, V., Erkut, E., Batta, R., 2000. On finding dissimilar paths. *European Journal of Operational Research*, 121(2), 232-246.
- Asakura, Y., Hato, E., 2004. Tracking survey for individual travel behaviour using mobile communication instruments. *Transportation Research Part C*, 12(3-4), 273-291.
- Azevedo, J.A., Santos Costa, M.E.O., Silvestre Madera, 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.
- Bar-Gera, H., 2007. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel. *Transportation Research Part C*, 15(6), 380-391.
- Basar, G., Bhat, C., 2004. A parameterized probabilistic consideration set model for airport choice: an application to the San Francisco Bay Area. *Transportation Research Part B*, 38(10), 889-904.
- Bekhor, S., Ben-Akiva, M.E., Ramming, S., 2002. Adaptation of logit kernel to route choice situation. *Transportation Research Record*, 1805, 78-85.
- 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.
- Bekhor, S., Prashker, J.N., 2001. Stochastic user equilibrium formulation for the generalized nested logit model. *Transportation Research Record*, 1752, 84-90.
- Bekhor, S., Prato, C.G., 2006. Effects of choice set composition in route choice modeling. *Proceedings of the 11th International Conference on Travel Behavior Research*, Kyoto, Japan.
- Bekhor, S., Prato, C.G., 2009. Methodological transferability in route choice modeling. *Transportation Research Part B*, 43(4), 422-437.
- Bekhor, S., Toledo, T., 2005. Investigating path-based solution algorithms to the stochastic user equilibrium problem. *Transportation Research Part B*, 39(3), 279-295.

- Bekhor, S., Toledo, T., Prashker, J.N., 2006. Implementation issues of route choice models in path-based algorithms. Proceedings of the 11th International Conference on Travel Behavior Research, Kyoto, Japan.
- Bell, M.G.H., 1995. Alternatives to Dial's logit assignment algorithm. Transportation Research Part B, 29(4), 287-295.
- Bellman, R., 1958. On a routing problem. Quarterly of Applied Mathematics, 16 (1), 87-90.
- Bellman, R., Kalaba, R., 1960. On  $k$ -th best policies. Journal of the Society for Industrial and Applied Mathematics, 8, 582-585.
- Ben-Akiva, M.E., Bergman, M.J., Daly, A.J., Ramaswamy, R., 1984. Modeling inter-urban route choice behaviour. In: Volmuller, J., Hamerslag, R. (Eds.), Proceedings of the 9th International Symposium on Transportation and Traffic Theory. VNU Science Press, Utrecht, The Netherlands, 299-330.
- Ben-Akiva, M.E., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. In: Hall, R.W. (Ed.), Handbook of Transportation Science. Kluwer, Dordrecht, The Netherlands, 5-34.
- Ben-Akiva, M.E., Boccara, B., 1995. Discrete choice models with latent choice sets. International Journal of Research in Marketing, 12, 9-24.
- Ben-Akiva, M.E., Bolduc, D., 1996. Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working Paper, Massachusetts Institute of Technology, Cambridge, USA.
- Ben-Akiva, M.E., Bolduc, D., Daly, A., 1993. Estimation of travel choice models with randomly distributed values of time. Transportation Research Record, 1413, 88-97.
- Benjamins, M., Lindveld, K., Van Nes, R., 2002. Multimodal travel choice modelling: a supernetwork approach. Proceedings of the 81th Annual Meeting of the Transportation Research Board, Washington, D.C.
- Bierlaire, M., Frejinger, E., 2005. Route choice models with subpath components. Proceedings of the 5th Swiss Transport Research Conference, Ascona, Switzerland.
- Bierlaire, M., Frejinger, E., 2008. Route choice modeling with network-free data. Transportation Research Part C, 16(2), 187-198.
- Bliemer, M.C.J., Bovy, P.H.L., 2008. Impact of route choice set on route choice probabilities. Transportation Research Record, 2076, 10-19.
- Bliemer, M.C.J., Bovy, P.H.L., Li, H., 2007. Some properties and implications of stochastically generated route choice sets. Proceedings of the 6th Tristan Conference, Phuket, Thailand.
- Bliemer, M.C.J., Taale, H., 2006. Route generation and dynamic traffic assignment for large networks. Proceedings of the First International Symposium on Dynamic Traffic Assignment, Leeds, UK, 90-99.
- Bovy, P.H.L., 2009. On modelling route choice sets in transportation networks: a synthesis. Transport Reviews, 29(1), 43-68.
- Bovy, P.H.L., Bekhor, S., Prato, C.G., 2008. The factor of revised path size: an alternative derivation. Transportation Research Record, 2076, 132-140.
- Bovy, P.H.L., Bekhor, S., Prato, C.G., 2009. Route sampling correction for stochastic route choice set generation. Proceedings of the 88th Annual Meeting of the Transportation Research Board, Washington, D.C.
- Bovy, P.H.L., Fiorenzo-Catalano, S., 2007. Stochastic route choice set generation: behavioral and probabilistic foundations. Transportmetrica, 3(3), 173-189.
- Cantillo, V., Heydecker, B., Ortuzar, J.d.D., 2006. A discrete choice model incorporating thresholds for perception in attribute values. Transportation Research Part B, 40(9), 807-825.

- Cantillo, V., Ortuzar, J.d.D., 2005. A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transportation Research Part B*, 39(7), 641-657.
- Cascetta, E., 2001. *Transportation Systems Engineering: Theory and Methods*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Lesort, J.B. (Ed.), *Proceedings of the Thirteenth International Symposium on Transportation and Traffic Theory*. Pergamon, Lyon, France, 697-711.
- Cascetta, E., Papola, A., 2001. Random utility models with implicit availability perception of choice travel for the simulation of travel demand. *Transportation Research Part C*, 9(4), 249-263.
- Cascetta, E., Russo, F., Vitetta, A., 1997. Stochastic user equilibrium assignment with explicit path enumeration: comparison of models and algorithms. In: Papageorgiou, M., Pouliezios, A. (Eds.), *Proceedings of 8th International Symposium on Transportation Systems*. Elsevier, Oxford, U. K, 1031-1037.
- Caussade, S., Ortuzar, J.d.D., Rizzi, L.I., Hensher, D.A., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B*, 39(7), 621-640.
- Chu, C., 1989. A paired combinatorial logit model for travel demand analysis. *Proceedings of the 5th World Conference on Transportation Research*, Ventura, USA, 295-309.
- Daganzo, C.F., Sheffi, Y., 1977. On stochastic models of traffic assignment. *Transportation Science*, 11, 253-274.
- De la Barra, T., Perez, B., Anez, J., 1993. Multidimensional path search and assignment. *Proceedings of the 21st PTRC Summer Annual Meeting*, Manchester, England, 307-319.
- Dial, R.B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5(2), 83-111.
- Dial, R.B., 1997. Bicriterion traffic assignment: efficient algorithms plus examples. *Transportation Research Part B*, 31(5), 357-379.
- Dial, R.B., 2006. A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. *Transportation Research Part B*, 40(10), 917-936.
- Dijkstra, E.W., 1959. A note on two problems in connection with graphs. *Numerical Mathematics*, 1, 269-271.
- Dreyfus, S.E., 1969. An appraisal of some shortest path algorithms. *Operational Research* 17, 395-412.
- Dougherty, M., 1995. A review of neural networks applied to transport. *Transportation Research Part C*, 3(4), 247-260.
- Eppstein, D., 1998. Finding the  $K$  shortest paths. *Journal of the Society for Industrial and Applied Mathematics*, 28(2), 652-673.
- Fiorenzo-Catalano, S., Van der Zij, N.J., 2001. A forecasting model for inland navigation based on route enumeration. *Proceedings of the European Transport Conference*, PTRC Education and Research Services Ltd., London, 1-11.
- Fisk, C., 1980. Some developments in equilibrium traffic assignment. *Transportation Research Part B* 14 (3), 243-255.



- Frejinger, E., 2007. Random sampling of alternatives in a route choice context. Proceedings of the European Transport Conference, Leeuwenhorst, The Netherlands.
- Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B*, 41(3), 363-378.
- Frejinger, E., Bierlaire, M., Ben-Akiva, M.E., 2009. Sampling of alternatives for route choice modeling. *Transportation Research Part B*, doi:10.1016/j.trb.2009.03.001 (in press).
- Freundschuh, S.M., 1992. Is there a relationship between spatial cognition and environmental patterns? *Lecture Notes in Computer Science*, 639, 288-304.
- Friedrich, M., Hofsäß, I., Wekeck, S., 2001. Timetable-based transit assignment using branch & bound. *Transportation Research Record*, 1752, 100-107.
- Gale, N., Golledge, R.G., Pellegrino, J.W., Doherty, S., 1990. The acquisition and integration of route knowledge in an unfamiliar neighbourhood. *Journal of Environmental Psychology*, 10, 3-25.
- Garling, T., Golledge, R.G., 2000. Cognitive mapping and spatial decision making. In Kitchin, R., Freundschuh, S.M. (Eds.), *Cognitive Mapping: Past, Present, and Future*. Routledge, London, 44-65.
- Gliebe, J.P., Koelman, F., Ziliaskopoulos, A., 1999. Route choice using a paired combinatorial logit model. Proceedings of the 78th Annual Meeting of the Transportation Research Board, Washington, D.C.
- Hadjiconstantinou, E., Christofides, N., 1999. An efficient implementation of an algorithm for finding k-shortest simple paths. *Networks*, 34, 88-101.
- Han, B., Algiers, S., Engelson, L., 2001. Accommodating drivers' taste variation and repeated choice correlation in route choice modeling by using the mixed logit model. Proceedings of the 80th Annual Meeting of the Transportation Research Board, Washington, D.C.
- Henn, V., 2000. Fuzzy route choice model for traffic assignment. *Fuzzy Sets and Systems*, 116(11), 77-101.
- Hess, S., Bierlaire, M., Polak, J., 2005. Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A*, 39(2-3), 221-236.
- Hess, S., Polak, J., Bierlaire, M., 2006. Confounding between taste heterogeneity and error structure in discrete choice models. Proceedings of the European Transport Conference, Strasbourg, France.
- Hoffman, W., Pavley, R., 1959. A method for the solution of the n-th best path problem. *Journal of the Association for Computing Machinery*, 6, 506-514.
- Hoogendoorn-Lanser, S., 2005. Modelling travel behaviour in multi-modal networks. Ph.D. Thesis, TRAIL Research School, Technical University of Delft, The Netherlands.
- Hoogendoorn-Lanser, S., Van Nes, R., Bovy, P.H.L., 2005. Path size and overlap in multi-modal transport networks: a new interpretation. In: Mahmassani H.S. (Ed.), *Transportation and Traffic Theory, Flow dynamics and Human Interaction*. Elsevier, Oxford, UK, 63-84.
- Horowitz, J.L., Louviere, J.J., 1995. What is the role of consideration sets in choice modeling? *International Journal of Research in Marketing*, 12, 39-54.
- Hunt, D.T., Kornhauser, A.L., 1997. Assigning traffic over essentially-least-cost paths. *Transportation Research Record*, 1556, 1-7.
- Jan, O., Horowitz, A., Peng, Z., 2000. Using GPS data to understand variations in path choice. *Transportation Research Record* 1706, 145-151.

- Jou, R.C., 2001. Modeling the impact of pre-trip information on commuter departure time and route choice. *Transportation Research Part B*, 35(10), 887-902.
- Katoh, N., Ibaraki, T., Mine, H., 1982. An efficient algorithm for k shortest simple paths. *Networks*, 12, 411-427.
- Koelman, F., Wen, C., 1998. Alternative nested logit models: structure, properties and estimation. *Transportation Research Part B*, 32(5), 289-298.
- Kuby, M., Zhongyi, X., Xiaodong, X., 1997. A minimax method for finding the k-best differentiated paths. *Geographical Analysis*, 29(4), 298-313.
- Lam, T.C., Small, K., 2001. The value of time and reliability: measurement from a value pricing experiment. *Transportation Research Part E*, 37(2-3), 231-251.
- Lawler, E.L., 1976. *Combinatorial Optimization: Networks and Matroids*. Holt, Rinehart & Winston, New York.
- Li, H., Guensler, R., Ogle, J., 2005. Analysis of morning commute route choice patterns using Global Positioning System-based vehicle activity data. *Transportation Research Record*, 1926, 162-170.
- Lombard, K., Church, R.L., 1993. The gateway shortest path problem: generating alternative routes for a corridor location problem. *Geographical Systems*, 1, 25-45.
- Lotan, T., 1997. Effects of familiarity on route choice behavior in the presence of information. *Transportation Research Part C*, 5(3), 225-243.
- Lotan, T., Koutsopoulos, H., 1993. Models for route choice behavior in the presence of information using concepts from fuzzy set theory and approximate reasoning. *Transportation*, 20(2), 129-155.
- Maher, M., 1998. Algorithms for logit-based stochastic user equilibrium assignment. *Transportation Research Part B*, 32(8), 539-549.
- Manski, C., 1977. The structure of random utility models. *Theory and Decision*, 8 (3), 229-254.
- McFadden, D., Train, T., 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447-470.
- Murakami, E., Wagner, D., 1999. Can global positioning system (GPS) improve trip reporting? *Transportation Research Part C*, 7(2-3), 149-165.
- Nakayama, S., Kitamura, R., 2000. Route choice model with inductive learning. *Transportation Research Record*, 1725, 63-70.
- Nakayama, S., Kitamura, R., Fujii, S., 2001. Drivers' route choice rules and network behavior: do drivers become rational and homogeneous through learning? *Transportation Research Record* 1752, 62-68.
- Nielsen, O.A., 2000. A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B*, 34(5), 377-402.
- Nielsen, O.A., 2004. Behavioural responses to pricing schemes: description of the Danish AKTA experiment. *Journal of Intelligent Transportation Systems*, 8(4), 233-251.
- Nielsen, O.A., Daly, A., Frederiksen, R.D., 2002. A stochastic multi-class road assignment model with distributed time and cost coefficients. *Networks and Spatial Economics*, 2, 327-346.
- Nielsen, O.A., Frederiksen, R.D., 2006. Optimisation of timetable-based, stochastic transit assignment models based on MSA. *Annals of Operations Research*, 144(1), 263-285.
- Park, D., Rilett, L.R., 1997. Identifying multiple and reasonable paths in transportation networks: a heuristic approach. *Transportation Research Record* 1607, 31-37.
- Prashker, J.N., Bekhor, S., 1998. Investigation of stochastic network loading procedures. *Transportation Research Record*, 1645, 94-102.

- Prashker, J.N., Bekhor, S., 2000. Congestion, stochastic, and similarity effects in stochastic user equilibrium models. *Transportation Research Record*, 1733, 80-87.
- Prashker, J.N., Bekhor, S., 2004. Route choice models used in the stochastic user equilibrium problem: a review. *Transport Reviews*, 24(4), 437-463.
- Prato, C.G., 2005. Latent factors and route choice behaviour. Ph.D. Thesis, Turin Polytechnic, Italy.
- Prato, C.G., Bekhor, S., 2006. Applying branch & bound technique to route choice set generation. *Transportation Research Record*, 1985, 19-28.
- Prato, C.G., Bekhor, S., 2007. Modeling route choice behavior: how relevant is the choice set composition? *Transportation Research Record*, 2003, 64-73.
- Ramming, S., 2002. Network knowledge and route choice. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, USA.
- Richardson, A., 1982. Search models and choice set generation. *Transportation Research Part A*, 16(5-6), 403-419.
- Ridwan, M., 2004. Fuzzy preference based traffic assignment problem. *Transportation Research Part C*, 12(3-4), 209-233.
- Rilett, L.R., Park, D., 2001. Incorporating uncertainty and multiple objectives in real-time route selection. *Journal of Transportation Engineering*, 127(6), 531-539.
- Ruphail, N.M., Ranjithan, S.R., ElDessouki, W., Smith, T., Brill, E.D., 1995. A decision support system for dynamic pre-trip route planning. Applications of advanced technologies. In: *Transportation Engineering: Proceedings of The Fourth International Conference*, 325-329.
- Scott, K., Pabon-Jimenez, G., Bernstein, D., 1997. Finding alternatives to the best path. Proceedings of the 76th Annual Meeting of the Transportation Research Board, Washington, D.C.
- Sheffi, Y., 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, USA.
- Sheffi, Y., Powell, W.B., 1982. An algorithm for the equilibrium assignment problem with random link times. *Networks*, 12, 191-207.
- Shier, D.R., 1976. Iterative methods for determining the K shortest paths in a network. *Networks*, 6, 205-229.
- Shier, D.R., 1979. On algorithms for finding the K-shortest paths in a network. *Networks*, 9, 195-214.
- Swait, J., 2001. Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B*, 35(7), 643-666.
- Swait, J., Ben-Akiva, M.E., 1987. Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B*, 21(2), 91-102.
- Train, K., Sonnier, G., 2005. Mixed logit with bounded distributions of correlated partworths. In: Scarpa, R., Alberini, A. (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics*. Springer, Dordrecht, 117-134.
- Van der Zij, N.J., Fiorenzo-Catalano, S., 2005. Path enumeration by finding the constrained K-shortest paths. *Transportation Research Part B*, 39(6), 545-563.
- Vovsha, P., 1997. The cross-nested logit model: application to mode choice in the Tel Aviv metropolitan area. *Transportation Research Record*, 1607, 13-20.
- Vrtic, M., Schussler, N., Erath, A., Axhausen, K., Frejinger, E., Bierlaire, M., Stojanovic, S., Rudel, R., Maggi, R., 2006. Including travelling costs in the modelling of mobility behaviour. Final report for SVI research program Mobility Pricing: Project B1, on behalf of the Swiss Federal Department of the Environment, Transport, Energy and Communications, IVT ETH Zurich, ROSO EPF Lausanne and USI Lugano.

- Wen, C., Koelman, F., 2001. The generalized nested logit model. *Transportation Research Part B*, 35(7), 627-641.
- Wolf, J., Schönfelder, S., Samaga, U., Oliveira, M., Axhausen, K.W., 2004. Eighty weeks of Global Positioning System traces: approaches to enriching trip information. *Transportation Research Record*, 1870, 46-54.
- Yai, T., Iwakura, S., Morichi, S., 1997. Multinomial probit with structured covariance for route choice behavior. *Transportation Research Part B*, 31(3), 195-207.
- Yamamoto, T., Kitamura, R., Fujii, J., 2002. Drivers' route choice behavior: analysis by data mining algorithms. *Transportation Research Record*, 1807, 59-66.
- Yang, H., Kitamura, R., Jovanis, P.P., Vaughn, K.M., Abdel-Aty, M. A., 1993. Exploration of route choice behavior with advanced traveler information using neural network concepts. *Transportation*, 20(2), 199-223.
- Yen, J.Y., 1971. Finding the K shortest loopless paths in a network. *Management Science*, 17, 712-716.
- Zhou, Z., Chen, A., 2003. Stochastic user equilibrium problem: a comparison between length-based and congestion-based C-Logit models. In: Loo, B.P.Y., Lam, S.W.K. (Eds.), *Proceedings of the 8th Hong Kong Society of Transportation Studies Conference: Transportation and Logistics*. Hong Kong, China, 244-253.