# A link based network route choice model with unrestricted choice set

Mogens Fosgerau [a,b], Emma Frejinger [c,*], Anders Karlstrom [d]

[a] Technical University of Denmark, Denmark
[b] Centre for Transport Studies, Sweden
[c] DIRO, Université de Montréal, Canada
[d] KTH Royal Institute of Technology, Dept. of Transport Science, Division of Transport and Location Analysis, and Centre for Transport Studies, Stockholm, Sweden

## ARTICLE INFO

## ABSTRACT

This paper considers the path choice problem, formulating and discussing an econometric random utility model for the choice of path in a network with no restriction on the choice set. Starting from a dynamic specification of link choices we show that it is equivalent to a static model of the multinomial logit form but with infinitely many alternatives. The model can be consistently estimated and used for prediction in a computationally efficient way. Similarly to the path size logit model, we propose an attribute called link size that corrects utilities of overlapping paths but that is link additive. The model is applied to data recording path choices in a network with more than 3000 nodes and 7000 links.

© 2013 Elsevier Ltd. All rights reserved.

## 1. Introduction

Consider a directed connected graph, defined in terms of links (arcs) and nodes (vertices). A number of deterministic attributes are associated with each link. A path is a sequence of links that connects an origin to a destination and path attributes are assumed to be link additive.[1] Given a sample of observations of origins, destinations and connecting paths, the objective is to formulate an econometric model for the choice of path conditional on origins and destinations. It is required that the model assigns probabilities to paths in a way that is consistent with rational behavior and that model parameters can be consistently estimated. This paper describes and analyzes such a model that we call recursive logit.

Estimated route choice models and the model parameters describing such choices have a variety of uses and the diffusion of GPS technology means that datasets containing directly observed path choices are becoming increasingly available. One application is in models predicting traffic load in road networks; such models are widely used to assess, e.g., infrastructure investments, regulation of traffic through road pricing, capacity management, etc. Another application is in route guidance software. These rely on well-known algorithms for finding the least cost path in a network (Dijkstra, 1959). However, it is necessary to be able to compute the cost of a path, when the path is described in terms of not just one but a range of characteristics such as length, speed, likelihood of delay, left turns, and speed bumps. Using a route choice model it is possible to estimate relevant parameters based on observed choices.

---

\* Corresponding author.
   E-mail addresses: mf@transport.dtu.dk (M. Fosgerau), emma.frejinger@umontreal.ca (E. Frejinger), amail@kth.se (A. Karlstrom).
   [1] Path and route are used interchangeably in the route choice modeling literature. In this paper we use path except when referring to a choice model because they are in general called route choice models in the literature.

The recursive logit model corresponds to a dynamic discrete choice model (Rust, 1987) where the path choice problem is formulated as a sequence of link choices. At each node the decision maker chooses the utility-maximizing outgoing link with link utilities given by the instantaneous cost, the expected maximum utility to the destination (value function) and i.i.d. extreme value type I error terms. Link choice probabilities are then given by a multinomial logit model and expected downstream utilities are identified from Bellman equations. This leads to a chosen path in a sequential and dynamic fashion. When the network contain cycles, paths may contain loops and can be arbitrarily long; then there will generally be infinitely many potential paths that connect an origin to a destination. It turns out that the probability that a given path is chosen has the form of a multinomial logit model, except that the number of alternatives is infinite.

So far, all route choice models that can be estimated from observed paths have been path based, meaning that they describe discrete choices among paths. The recursive logit has the advantage over these approaches that it can be both consistently estimated and used for prediction in a computationally efficient way without requiring any restriction on the choice sets of paths. (We further discuss the comparison with existing models in Section 7.) This is potentially very useful for e.g. traffic simulation applications where the number of paths that can be stored is restricted by available memory.

The idea of using a sequential link choice model to describe path choice has been around for quite some time in the context of traffic assignment (e.g. Bell, 1995; Akamatsu, 1996; Baillon and Cominetti, 2008). There is also literature in the context of networks without cycles (e.g. Dial, 1971), which we do not review here because of this restriction on the choice set. Melo (2012) presents a sequential choice model similar to ours but where a path is interpreted as a bundle of goods (this work has been developed simultaneously and independently of ours) and the main result is a Markovian assignment which is related to the work by Baillon and Cominetti (2008).

This paper is the first to provide links between the sequential link based route choice model and the finite multinomial logit model in the context of route choice. Second, we are also the first to provide an interpretation of the model as a dynamic discrete choice model (Rust, 1987; Aguirregabiria and Mira, 2010), connecting the path choice problem to that literature. Third, we are the first to discuss estimation of the model, including by sampling of alternatives (McFadden, 1978).

Before formalizing the recursive logit model in the following section, we briefly introduce the multinomial logit model for a finite set of alternatives $j \in \{1, \ldots, J\}$ as a basis for comparison. A utility $u_j$ is associated with each alternative and is the sum of a deterministic and a random component $v_j + \mu \varepsilon_j$ where $\varepsilon_j$ are i.i.d. extreme value type 1 with scale parameter $\mu$. The maximum utility is $u_{\max} = \max_j u_j$ and the expected maximum utility is $Eu_{\max} = \mu \ln \sum_j e^{\frac{1}{\mu} v_j}$. It is a general fact for additive random utility models (McFadden, 1978; Fosgerau et al., 2013) that choice probabilities can be found as the gradient of $Eu_{\max}$ considered as a function of the vector of deterministic utility components $v$ and hence

$$P_j = \frac{e^{\frac{1}{\mu} v_j}}{e^{\frac{1}{\mu} Eu_{\max}}} = \frac{e^{\frac{1}{\mu} v_j}}{\sum_{j'} e^{\frac{1}{\mu} v_{j'}}}. \tag{1}$$

The multinomial logit model satisfies the IIA property, since $P_i/P_j = e^{\frac{1}{\mu}(v_i - v_j)}$, which depends only on $v_i - v_j$. This property makes it possible to estimate the model using only a sample of alternatives in order to reduce computational cost (McFadden, 1978). We describe in Section 3.4 how this can be done in the context of the recursive logit model.

The proposed model has an infinite but still discrete choice set. It is worth mentioning that there is a small literature that considers generalizations of discrete choice models to continuous choice sets representing physical space or time. The first to describe such models seems to be McFadden (1976).

Section 2 introduces the recursive logit model; Section 3 discusses various properties of the model including computation of the likelihood and estimation using sampling of alternatives. Section 4 introduces the link size attribute and presents an illustrative example. Maximum likelihood estimation of the model is described in Section 5 and 6 provides estimation results using synthetic and real data. In Section 7 we discuss comparison with other route choice models and finally Section 8 concludes.

## 2. The recursive logit model

In this section we formulate the path choice problem as a dynamic discrete choice model where the utility maximization problem is consistent with a dynamic programming problem. For the sake of simplicity and because the numerical results in this paper concern a static network we do not time index the notation. The model formulation is however valid for a dynamic setting as long as the link attributes are deterministic.

Consider a directed connected graph (not assumed acyclic) $G = (A, \mathcal{V})$ where $A$ is the set of links and $\mathcal{V}$ the set of nodes. We denote links $k, a \in A$ and the set of outgoing links from the sink node of $k$ $A(k)$ (see Fig. 1 for an illustration of the notation). A
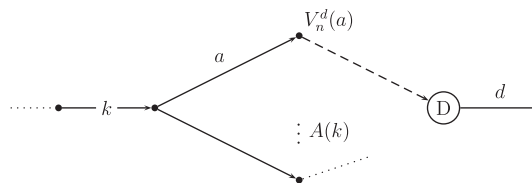


**Fig. 1.** Illustration of notation.

path is a sequence of links $(k_0, \ldots, k_I)$ with $k_{i+1} \in A(k_i)$ for all $i < I$. A deterministic utility component $v_n(a|k) = v(x_{n,a|k}; \beta) < 0$ is associated with each link pair, $a \in A(k) \subset A$, where $x_{n,a|k}$ is a vector of observed characteristics of the link pair $(k, a)$ that may include characteristics of traveler $n$ and $\beta$ is an unknown parameter vector to be estimated. Path attributes are assumed to be link additive. In the terminology of dynamic programming, $k$ is a state and $a$ is a potential action given state $k$.

We note that the choice of state space depends on the application. For instance, if a network has time-dependent link travel times, a state can be defined by a link time pair and the size of the state space is then the number of links times the number of time intervals. In some cases it may be possible to increase the state space in order to include attributes that would otherwise not be link additive, similar to what is done here. Indeed, dealing with link pairs instead of node-link pairs allows us to include turn related attributes in $x_{n,a|k}$.

Consider now an individual traveling from an origin to a destination node. The network is extended with a dummy link to include the destination. More precisely, we define an absorbing state by adding a link $d$ without successors from the destination node (see Fig. 1). The set of all links is then $\widetilde{A}^d = A \cup d$. The deterministic utility is $v(d|k) = 0$ for all $k$ that have the destination as sink node.

At each state $k$ a traveler chooses an action $a$ (next link) from the set of outgoing links $A(k)$. An instantaneous utility, $u_n(a|k) = v_n(a|k) + \mu \varepsilon_n(a)$, is associated with each action in the choice set $A(k)$ conditional on current state $k$. The random terms $\varepsilon_n(a)$ are assumed i.i.d. extreme value type 1 with zero mean and they are independent of everything else in the model.

The traveler chooses the next link given the current state in a stochastic process having the Markov property (Rust, 1987; Aguirregabiria and Mira, 2010). In our setting the next state is just the chosen link and it is given with certainty by the action. At each current state $k$ the traveler observes random utility terms $\varepsilon_n(a)$, $a \in A(k)$. He then chooses the link $a$ that maximizes the sum of instantaneous utility $u_n(a|k)$ and expected downstream utility, the latter being given by the value function $V^d(a)$, which is defined by taking the continuation of this process into account via the Bellman equation (Bellman, 1957) as follows.

$$V_n^d(k) = E\left[ \max_{a \in A(k)} \left( v_n(a|k) + V_n^d(a) + \mu \varepsilon_n(a) \right) \right] \quad \forall k \in A. \tag{2}$$

Note here that the traveler may traverse a link more than once and that new random terms are drawn at each step.

The probability of choosing a link $a$ given state $k$ is given by the multinomial logit model

$$P_n^d(a|k) = \frac{e^{\frac{1}{\mu}(v_n(a|k) + V_n^d(a))}}{\sum_{a' \in A(k)} e^{\frac{1}{\mu}(v_n(a'|k) + V_n^d(a'))}}. \tag{3}$$

Then the value function is the logsum

$$V_n^d(k) = \begin{cases} \mu \ln \sum_{a \in A} \delta(a|k) e^{\frac{1}{\mu}(v_n(a|k) + V_n^d(a))} & \forall k \in A \\ 0 & k = d. \end{cases} \tag{4}$$

For the purpose of introducing matrix notation in the next section, we have added an indicator $\delta(a|k)$ that equals one if $a \in A(k)$ and zero otherwise so that the sum is over the set $A$. $V(d) = 0$ since the maximum over an empty set is taken to be zero.

## 3. Properties of the model

In this section we present how to solve the Bellman Eq. (2) and how to compute path choice probabilities and link flows. Furthermore, we discuss sampling of alternatives. In order to simplify the notation, we suppress the index $n$ for individuals and the superscript $d$ used previously to clarify that value functions and link choice probabilities are destination specific.

### 3.1. Solving the Bellman equations

We start by noting that our formulation has some features that are non-standard in the dynamic discrete choice literature. First, although we formulate the problem as an infinite horizon problem, we have a discount factor of one. Second, we assume a deterministic environment, such that a given state and action will result in a new state deterministically. This is an important, but realistic, assumption in our context, and as it is shown in the following, it allows us to compute the value functions by simply solving a systems of linear equations.

We transform (4) by taking the exponential and raising to the power $\frac{1}{\mu}$

$$e^{\frac{1}{\mu}V(k)} = \begin{cases} \sum_{a \in A} \delta(a|k) e^{\frac{1}{\mu}(v(a|k) + V(a))} & \forall k \in A, \\ 1 & k = d. \end{cases} \tag{5}$$

Eq. (5) can now be written in matrix form using the following definitions. Let $\mathbf{M}$ ($|\widetilde{A}| \times |\widetilde{A}|$) denote the incidence matrix defining instantaneous utilities

$$M_{ka} = \begin{cases} \delta(a|k)e^{\frac{1}{\mu}v(a|k)}, & a \in A(k) \\ 0 & \text{otherwise}. \end{cases} \tag{6}$$

Since $d$ has no successors, $\mathbf{M}$ has a zero row for $k = d$. We denote by $\mathbf{z}$ ($|\tilde{A}| \times 1$) a vector with elements $z_k = e^{\frac{1}{\mu}V(k)}$ and $\mathbf{b}$ ($|\tilde{A}| \times 1$) a vector with elements $b_k = 0$, $k \neq d$ and $b_d = 1$. Then (5) can be written as a system of linear equations

$$\mathbf{z} = \mathbf{Mz} + \mathbf{b} \iff (\mathbf{I} - \mathbf{M})\mathbf{z} = \mathbf{b}, \tag{7}$$

where $\mathbf{I}$ is the identity matrix. The system has a solution if $\mathbf{I} - \mathbf{M}$ is invertible, which may or may not be the case. This depends on the balance between the number of paths that connect the nodes in the network and the size of the instantaneous utilities $\frac{1}{\mu}v(a|k)$. This issue is particularly important to consider when searching over the parameter space in the estimation of the model since for some trial values of $\beta$, $\mathbf{I} - \mathbf{M}$ may be ill-conditioned or even singular. To explain this further, let $m \geqslant 0$ be some integer. The matrix $\mathbf{M}^m$ (the $m$th power of $\mathbf{M}$) then contains the instantaneous utilities of paths between any pair of links that are $m$ links long. On the one hand, many alternative paths and values of $\frac{1}{\mu}v(a|k)$ close to zero tend to make the entries of $\mathbf{M}^m$ large. That is, the expected utility of traveling between two locations in $m$ steps is larger the more ways there are of doing that trip. If there are enough alternative paths and if the instantaneous utilities are sufficiently close zero, then the expected utility may be positive. Indeed, the deterministic utilities are negative by definition but the random terms may be positive. On the other hand, if the elements of $\mathbf{M}^m$ tend to 0 as $m \to \infty$ then $\mathbf{I} - \mathbf{M}$ is invertible. To conclude this discussion on the invertibility of $\mathbf{I} - \mathbf{M}$ we note that $\frac{1}{\mu}v(a|k)$ is network dependent since they are scaled to the variance of the error terms. Therefore, dense networks and many alternative paths do not necessarily imply that $\mathbf{I} - \mathbf{M}$ is ill-conditioned.

For medium size networks the system (7) can be solved using a direct solution method. For the numerical results in this paper, reported in Section 6, we use the direct solver available in MATLAB and our network has 7459 links. If the network is very large a direct method may not be feasible and one can then use iterative solution methods, see e.g. Saad and vander Vorst (2000) for an excellent review of such methods.

Note now that the next-link choice probabilities for a common destination are independent of the origin. So, we organize the probabilities (3) into a matrix $\mathbf{P}$ defined for links in $A$. A row corresponding to state $k$ is

$$\mathbf{P}_k = \frac{\mathbf{M}_k \circ \mathbf{z}^T}{\mathbf{M}_k \mathbf{z}} \tag{8}$$

where $\circ$ is the element-by-element product and $\mathbf{M}_k$ is row $k$ of matrix $\mathbf{M}$.

### 3.2. Path probabilities

A path observation is a sequence of states $\sigma = \{k_i\}_{i=0}^I$ where $k_0$ is the origin and $k_I = d$. The likelihood of a path observation is $P(\sigma) = \prod_{i=0}^{I-1} P(k_{i+1}|k_i)$ by the Markov property of the model. A simple expression for $P(\sigma)$ is available using (3) upon noting that the denominator is just $e^{\frac{1}{\mu}V(k)}$. Hence, since $V(d) = 0$ by definition,

$$P(\sigma) = \prod_{i=0}^{I-1} e^{\frac{1}{\mu}(v(k_{i+1}|k_i) + V(k_{i+1}) - V(k_i))} = e^{-\frac{1}{\mu}V(k_0)} \prod_{i=0}^{I-1} e^{\frac{1}{\mu}v(k_{i+1}|k_i)}. \tag{9}$$

Denote $v(\sigma) = \sum_{i=0}^{I-1} v(k_{i+1}|k_i)$, such that

$$P(\sigma) = \frac{e^{\frac{1}{\mu}v(\sigma)}}{e^{\frac{1}{\mu}V(k_0)}} = \frac{e^{\frac{1}{\mu}v(\sigma)}}{\sum_{\sigma' \in \Omega} e^{\frac{1}{\mu}v(\sigma')}} \tag{10}$$

where $\Omega$ is the set of all possible paths, which is infinite in our case.

Note here the similarity to the multinomial logit model as written in (1). The numerator in (10) comprises the deterministic utility component for the path $\sigma$ and the denominator comprises the expected utility for the choice of path; the denominator equals the sum of $e^{\frac{1}{\mu}v(\sigma')}$ over all possible paths. Hence our model is equivalent to a path based multinomial logit model with an infinite number of alternatives. The ratio of the probabilities of any two paths $\sigma_1$ and $\sigma_2$ connecting the same origin and destination depends only on $v(\sigma_1) - v(\sigma_2)$, this is the IIA property.

### 3.3. Link flows

In some situations it is useful to be able to apply the model to predict flows on the links in the network. The model allows a simple way of doing this using the next-link probabilities (3). We compute the link flows for one destination and multiple origins. The link flows for multiple destinations can then be computed by summation.

Consider a destination $d$. Denote the demand for trips originating at link $a$ and ending at $d$ as $G(a)$. Denote the expected flow on link $a$ as $F(a)$. This comprises the flow that originates on $a$ and the expected incoming flow, that is, $F(a) = G(a) + \sum_{k \in A} P(a|k)F(k)$. We organize the expected flows into a vector $\mathbf{F}$, the demand into $\mathbf{G}$, both defined for $A$.

Using this notation we have

$$(\mathbf{I} - \mathbf{P}^T)\mathbf{F} = \mathbf{G}. \tag{11}$$

Baillon and Cominetti (2008) prove that $(\mathbf{I} - \mathbf{P}^T)$ is invertible (Lemma 1) and we can therefore compute the expected link flows by solving the system of linear equations.

### 3.4. Estimation with sampling of alternatives

Even though the focus of this paper is the specification of a model that can be consistently estimated and applied without sampling choice sets of paths, this section is dedicated to linking the proposed model to the path sampling approach (Frejinger et al., 2009). Indeed, the underlying choice set assumption is the same, namely, that any path in the network could be chosen. Moreover, showing that the results in McFadden (1978) for the logit model with a finite choice set also hold for the case with an infinite choice set is useful to understand the model.

A possible motivation for estimating the recursive logit model on a sample of alternatives is that a standard estimation software could be used. In this case the systems of linear Eq. (7) does not need to be solved in the iterations of the non-linear optimization algorithm. We can also note that the recursive logit model is very easy to use for sampling paths with known probability using (8). Hence, it could be used to sample paths and then estimate the model with a standard software using the sampling correction proposed in Frejinger et al. (2009).

Consider that for each observation, a sample of alternatives is chosen conditional on the observed path and that the number of sampled alternatives is bounded. Denote the probability of sampling a set of alternatives $D$ given observed choice $\sigma$ by $\pi(D|\sigma)$ and assume that the positive conditioning property holds: $\sigma \in D \Rightarrow \pi(D|\sigma) > 0$. We also assume that the chosen path is always included in $D$, such that $\sigma \notin D \Rightarrow \pi(D|\sigma) = 0$.

Then an argument from McFadden (1978) establishes the consistency of estimates based on the modified probabilities

$$\frac{e^{\frac{1}{\mu}v(\sigma)+\ln\pi(D|\sigma)}}{\sum_{\sigma'\in D}e^{\frac{1}{\mu}v(\sigma')+\ln\pi(D|\sigma')}}.$$

The advantage of this finding is that the modified likelihood is just that of a standard multinomial logit model having a finite number of alternatives.

The essential step in the McFadden (1978) argument, translated into the present context, is the following. The expected modified likelihood is

$$E\left(\sum_{\sigma\in\Omega}\sum_{D\subseteq\Omega}P^*(\sigma)\pi(D|\sigma)\ln\left[\frac{e^{\frac{1}{\mu}v(\sigma|\beta)+\ln\pi(D|\sigma)}}{\sum_{\sigma'\in\Omega}e^{\frac{1}{\mu}v(\sigma'|\beta)+\ln\pi(D|\sigma')}}\right]\right),$$

where the outer expectation is over explanatory variables including observation specific origins and destinations. This can be rewritten as

$$E\left(\sum_{D\subseteq\Omega}\left(\frac{\sum_{\sigma'\in\Omega}e^{\frac{1}{\mu}v(\sigma'|\beta^*)}\pi(D|\sigma')}{\sum_{\sigma'\in\Omega}e^{\frac{1}{\mu}v(\sigma'|\beta^*)}}\right)\sum_{\sigma\in D}\frac{e^{\frac{1}{\mu}v(\sigma|\beta^*)}\pi(D|\sigma)}{\sum_{\sigma'\in D}e^{\frac{1}{\mu}v(\sigma'|\beta^*)}\pi(D|\sigma')}\ln\left[\frac{e^{\frac{1}{\mu}v(\sigma|\beta)+\ln\pi(D|\sigma)}}{\sum_{\sigma'\in D}e^{\frac{1}{\mu}v(\sigma'|\beta)+\ln\pi(D|\sigma')}}\right]\right) \quad (12)$$

This may be recognized as the expectation of a term having the general form $E[\sum_{\sigma\in D}\phi(\sigma|\beta^*)\ln\phi(\sigma|\beta)]$, which is exactly the case considered by McFadden (1978). Consistency of the estimates then follows.

## 4. Link size attribute and illustrative example

In real networks, paths connecting a given origin destination pair share links. Due to this physical overlap it is generally thought that paths share unobserved attributes meaning that the path utilities are correlated. Ignoring this correlation may result in erroneous path probabilities and substitution patterns. Various models have been proposed in the literature to model correlated paths utilities. The most popular are the path size logit (Ben-Akiva and Bierlaire, 1999) and c-logit (Cascetta et al., 1996) models because they provide a way to heuristically correct the utility for overlapping paths while keeping the simple structure of a logit model. Several different versions of the path size (PS) attribute and commonality factor (the correction used in the c-logit model) have been proposed. They decrease the utility for overlapping paths by an amount that is somewhat proportional to the amount of overlap with other paths in the choice set. Hence, these corrections either require an enumeration of all paths or choice set sampling (see Frejinger et al. (2009) for an example). These corrections are not link additive and can therefore not be used with the recursive logit (RL) model. In this section we propose a link additive correction attribute, called Link Size (LS), that can be used with the RL model and that does not require any restriction on the choice set.

Path size logit (PSL) seems to be the most frequently used model in the literature and we therefore compare the LS attribute to the original formulation of the PS attribute. The latter is defined for a given path $j$ and choice set $C_n$

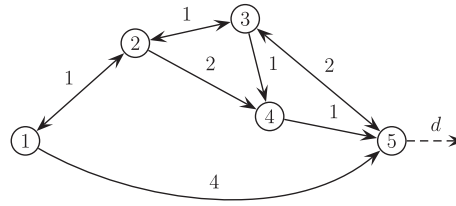$$PS_{jn} = \sum_{a\in\Gamma_j}\frac{L_a}{L_j}\frac{1}{\sum_{i\in C_n}\delta_{ai}} \quad (13)$$

**Fig. 2.** Illustrative example.

**Table 1**
Path probabilities for illustrative example.

| Paths | Logit | RL | PSL | RL with LS |
|---|---|---|---|---|
| 1: 12, 23, 35 | 0.25 | 0.2495 | 0.20 | 0.19 |
| 2: 12, 23, 34, 45 | 0.25 | 0.2502 | 0.12 | 0.13 |
| 3: 12, 24, 45 | 0.25 | 0.2495 | 0.20 | 0.19 |
| 4: 15 | 0.25 | 0.2495 | 0.48 | 0.48 |

where the sum is over all links of path $j$ (the set is denoted $\Gamma_j$). The length of a link $a$, $L_a$, is divided by the length of the path $L_j$ and the number of paths in choice set $C_n$ that uses link $a$ ($\delta_{ai}$ equals one if path $i$ uses link $a$ and zero otherwise). Therefore, $PS_{jn} = 1$ if path $j$ does not overlap with any other path in $C_n$ and $0 < PS_{jn} < 1$ otherwise. The correction of the path utility is $\beta_{PS} S \ln PS_{j}n$ where $\beta_{PS}$ is a parameter to be estimated. Note that if path $j$ does not overlap with any other path in $C_n$, $\ln PS_{j}n = 0$.

The LS attribute is based on a similar idea but instead of the number of paths using a given link, we use the expected link flow as a proxy for the amount of overlap. Indeed, we can apply the RL model to compute expected link flows in an efficient way using (11). To do so we choose an instantaneous utility with parameters $\widetilde{\beta}$, and we define an origin specific demand vector $\mathbf{G}^o$ with zero-valued elements except for the origin link which equals one. Note that we use $\widetilde{\beta}$ to denote when the parameters $\beta$ have chosen values, and later on we denote estimated values by $\widehat{\beta}$. The origin–destination specific vector $\mathbf{LS}^{od}$ of size $(A \times 1)$ where each element $LS_k^{od}$ is the LS attribute of link $k$ is

$$\mathbf{LS}^{od} = \mathbf{F}^{od}(\widetilde{\beta}). \tag{14}$$

$\mathbf{F}^{od}(\widetilde{\beta})$ is the solution to (11) which can be written for this specific case as $\left(\mathbf{I} - \mathbf{P}^d(\widetilde{\beta})^T\right)\mathbf{F}^{od}(\widetilde{\beta}) = \mathbf{G}^o$. Note that $\mathbf{F}^{od}$ may be viewed as normalized since $\mathbf{G}^o\mathbf{1} = 1$.

We use the small network shown in Fig. 2 as an illustrative example. Each bi-directional link actually corresponds to two directed links (arcs) and we denote links by the numbers of their source and sink nodes. Node 1 is the origin for the path based models and link 21 the origin for the RL model. We fix the cost of link 21 to zero so that the two specifications of the origin are equivalent. The destination node is 5 and a dummy link $d$ (absorbing state) is added for the RL model. The only attribute is link length and the values are given in the figure. Moreover, u-turns are not allowed.[2]

There are four simple paths (without loops), three going left at node 1 and one going right and they all have the same length of 4. In order to compute path probabilities we choose a length parameter $\widetilde{\beta}_l = -1.5$. In Table 1 we report the probabilities given by the logit model with the four simple paths (with and without the PS variable), and the RL model (with and without the LS attribute).

The logit model assigns the same uniform probability to each of the paths (0.25) and the probabilities of the RL model are numerically close. The difference in value is due to the fact that RL allows for loops with a very small probability. Of course, if we would increase the magnitude of $\widetilde{\beta}_l$ the probabilities tend to 0.25. The probabilities given by the PSL model using $\widetilde{\beta}_{PS} = 2.5$ are close to what often is expected in this kind of example, namely that there is roughly a 0.5 probability of going left at node 1 and 0.5 probability of going right.[3] Approximately the same result can be obtained applying the RL model with a LS attribute and $\widehat{\beta}_{LS} = -0.75$ (this attribute has been generated using the RL model with a length parameter $\widetilde{\beta}_l^{LS} = -1.5$). Since the LS and PS attributes are different corrections, it is clear that different parameter values need to be used. In this small example we choose values but in practice they should be estimated based on data.

The LS attribute depends on the chosen utility specification of the RL model used to compute the expected flow vector. However unlike the PS attribute no path sampling is required. The utility specification is application specific and different specifications should be tested to investigate the sensitivity of the final estimation results with respect to the definition of the LS attribute. Fig. 3 shows the probability of choosing one of the upper three paths in the example (going left at node 1). At least for this small example the probability is not very sensitive to the value of the length parameter $\widetilde{\beta}_l^{LS}$ used to

---

[2] Technically, this is implemented by introducing a fixed large penalty to each u-turn so that their probability is close to zero in the RL model, except for the origin link 21 to link 12.

[3] In reality one would expect that the possibility of adapting the route choice downstream is valuable and hence the probability of going left is higher than going right. In this example, the network is static and deterministic and only link length is considered so being adaptive has no value.
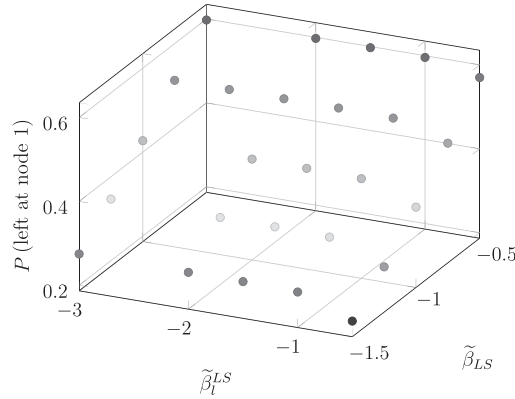
**Fig. 3.** Probability of going left at node 1 as a function of $\widetilde{\beta}_{LS}$ and $\widetilde{\beta}_l^{LS}$.

generate the LS attribute. As expected the probability of the correlated paths decrease as the magnitude of $\widetilde{\beta}_{LS}$ increase. Finally we note that including a LS attribute makes the RL model computationally more intensive to estimate and to apply because the matrix **M** becomes origin–destination specific.

## 5. Maximum likelihood estimation

The proposed model using the path probabilities as defined by (9) can be estimated by maximum likelihood. Since we are able to compute the value functions in a straightforward way we use an approach similar to Rust's nested fixed point algorithm.[4] Our estimation algorithm is implemented in MATLAB (code is available upon request) and we use the unconstrained non-linear optimization algorithm BFGS to search over the parameter space. For each trial value of the parameters, and destination (or origin–destination pair if a LS attribute is included) in the sample, the system of linear Eq. (7) is solved.

We note that the optimization problem is actually constrained since (7) does not have a solution for all possible values of the parameters and if there is no solution for at least one observation in the sample, then the log-likelihood function is not defined (see Section 2 for a discussion on the invertibility of **I** − **M**). We deal with this issue by starting at a feasible point (large enough magnitude of the parameters) and then being conservative in the initial step size of the line search algorithm, at the price of an increased number of iterations. We hence make the innocuous assumption that infinite loops are not present in observed paths and that the specified model is sufficiently close to the true model so that (7) has a solution for all observations at the maximum likelihood estimates of the parameters.

The gradient is evaluated by finite difference at each iteration of BFGS. However, in order to increase the accuracy we use the analytical gradient for computing the standard errors. The log likelihood function defined for observations $n = 1, \ldots, N$ is $LL(\beta) = \ln \prod_{n=1}^{N} P(\sigma_n) = \frac{1}{\mu} \sum_{n=1}^{N} \sum_{i=0}^{I_n-1} v(k_{i+1}|k_i) - V(k_0)$ and if the instantaneous utilities are linear-in-parameters, the analytical derivative with respect to a given parameter $\beta_q$ is

$$\frac{\partial LL(\boldsymbol{\beta})}{\partial \beta_q} = \frac{1}{\mu} \sum_{n=1}^{N} \left[ \sum_{i=0}^{I_n-1} x_{i+1|i}^q - \frac{\partial V(k_0^n)}{\partial \beta_q} \right]. \tag{15}$$

To derive $\frac{\partial V(k_0^n)}{\partial \beta_q}$, we differentiate (7) which yields

$$\frac{\partial \mathbf{z}}{\partial \beta_q} = (\mathbf{I} - \mathbf{M})^{-1} \left[ \frac{\partial \mathbf{M}}{\partial \beta_q} \right] \mathbf{z}. \tag{16}$$

Since $z_k = e^{\frac{1}{\mu} V(k)}, \frac{\partial \mathbf{z}}{\partial \beta_q} = \mathbf{z} \circ \frac{1}{\mu} \frac{\partial \mathbf{V}}{\partial \beta}$ and accordingly

$$\frac{\partial V(k_0^n)}{\partial \beta_q} = \mu \frac{1}{z_0} \left( \frac{\partial \mathbf{z}}{\partial \beta_q} \right)_0 \tag{17}$$

where $z_0$ and $\left( \frac{\partial \mathbf{z}}{\partial \beta} \right)_0$ are elements corresponding to the origin. The analytical gradient could be used at each iteration of BFGS instead of the numerical gradient. Computing (15) is however more involved than computing the log-likelihood function $2|\beta|$ times (gradient by finite difference) since the log-likelihood function can be evaluated very efficiently thanks to (7) and (9).

---

[4] Another option is, e.g., the swapped nested fixed point algorithm proposed by Aguirregabiria and Mira (2002).

## 6. Numerical results

In this section we present estimation results using the network of Borlänge, a city located in the middle of Sweden. The network contains 3077 nodes and 7459 links with 21,452 link pairs (number of non-zero entries in the **M** matrix). First we present validation results using simulated observations followed by estimation results based on real GPS data.

### 6.1. Simulated observations

In order to validate the RL model we choose a model specification and use it to simulate 10 samples of 500 path observations each for one origin–destination pair in the network. The instantaneous utility function is defined by the following expression where we have chosen the parameters values ($\tilde{\beta}_{TT} = -2$, $\tilde{\beta}_{LT} = -1$, $\tilde{\beta}_{LC} = -1$ and $\tilde{\beta}_{UT} = -20$)

$$v(a|k) = \beta_{TT}TT_a + \beta_{LT}LT_{a|k} + \beta_{LC}LC_a + \beta_{UT}UT_{a|k}. \tag{18}$$

$TT_a$ is the travel time in minutes of link $a$ and $LT_{a|k}$ is a left turn dummy that equals one if the turn from $k$ to $a$ is a left turn with an angle larger than $40°$ (but less than $177°$) compared to the direction of the state link, and zero otherwise. $LC_a$ is a constant equal to one for all links which it is included to penalize paths with many crossings. $UT_{a|k}$ is a u-turn dummy that equals one if going from $a$ to $k$ is a turn of $177°$ or larger. In order to prevent numerical problems we have removed u-turns from the network.[5] If there are very short links in the network it can otherwise be possible to loop between a pair of nodes at almost no cost. The breakpoint for the u-turn angle is fixed so that small numerical differences do not matter (turn angle is 3.1 radians instead of pi). One third of the link pairs are classified as u-turns. With the left turn breakpoint angle fixed to $40°$, about one fifth of the link pairs are classified as left turns and this number is not very sensitive to this limit, for $30°$ it is still approximately one fifth.

Before presenting the estimation results we give some descriptive statistics of the 10 samples of 500 observations each. Taking the average over the samples, the observations use 39 different paths (with a maximum of 44 and minimum of 35) for this one origin–destination pair that is considered. The path travel times range from 8.8 min to 12.3 min with an average of 9 min. Out of the 5000 observations, 35 (0.7%) are paths with loops where we consider a repeated link as a loop. Indeed, as discussed previously, the model allows paths with loops with a very small probability. Out of these 35 observations, one had 4 loops, 8 had 2 loops and the remaining 26 observations had 1 loop.

Table 2 reports the parameter estimates and standard errors for each of the samples. All estimates are not significantly different from their true values at the 5% significance level. The last two rows in the table report the averages of the parameter estimates and of the estimated standard error as well as the standard error of the parameter estimates over the different samples. We find that the parameters and standard errors appear to be unbiased. Indeed, the average parameter estimates are close to the chosen parameter values and that of the standard error of the estimates are close to the average of the estimated standard errors.

### 6.2. Real observations

GPS traces of 200 vehicles were collected during two years for a traffic safety study in Borlänge with the main aim of monitoring speeds. After data processing, using a subset of 24 of the vehicles for which the log frequency of the GPS device was the highest, the sample consists of 1832 trips corresponding to simple paths with a minimum of five links. There are 466 destinations and more than 37,000 link choices in this sample.

The estimation results based on this data are presented in Table 3. The same instantaneous utility specification (18) as for the synthetic data is used for the RL model and the parameter associated with u-turns is fixed to $-20$ for numerical reasons, as explained in the previous section. The estimates are highly significant and the signs and magnitudes are reasonable. For the second model, we have included a LS attribute and estimated the associated parameter $\beta_{LS}$. The attribute is defined by (14) using the same utility specification as the RL model (18) with parameters $\tilde{\beta}_{TT} = -2.5$, $\tilde{\beta}_{LT} = -1$, $\tilde{\beta}_{LC} = -0.4$, $\tilde{\beta}_{UT} = -20$. There is a remarkable improvement in final log-likelihood value when adding the LS attribute and $\hat{\beta}_{LS}$ has its expected sign. Interpreting the parameters of the RL model with LS attribute in terms of travel time, a crossing (link constant) is worth approximately 6 s and a left turn approximately 20 s. Most intersections in the Borlänge network are uncontrolled and a left turn therefore does not only take longer time but also means crossing meeting traffic.

## 7. Discussion on route choice model comparison

There are basically three different route choice modeling approaches available in the literature including the one proposed here. The first approach, which we refer to as the classic approach is path based and uses choice sets which are the union the observed paths and those that have been sampled with some path generation algorithm. When estimating the model and using it for prediction the generated choice sets are treated as if they were the actual choice sets and no correction

---

[5] This has been implemented by choosing $\tilde{\beta}_{UT} = -20$ so that u-turns have a probability very close to zero. We have tested different values of $\beta_{UT}$ and conclude that the parameter estimates are insensitive to $\beta_{UT}$ for a range around $-20$.

**Table 2**
Simulation study results.

| Sample | $\widehat{\beta}_{TT}$ | Std. err. | $\widehat{\beta}_{LT}$ | Std. err. | $\widehat{\beta}_{LC}$ | Std. err. |
|---|---|---|---|---|---|---|
| 1 | −1.91 | 0.21 | −1.02 | 0.09 | −1.07 | 0.06 |
| 2 | −1.97 | 0.22 | −0.99 | 0.09 | −1.04 | 0.06 |
| 3 | −1.80 | 0.21 | −1.09 | 0.09 | −1.07 | 0.06 |
| 4 | −2.38 | 0.26 | −0.88 | 0.09 | −1.01 | 0.06 |
| 5 | −2.20 | 0.24 | −0.96 | 0.08 | −0.93 | 0.05 |
| 6 | −2.30 | 0.26 | −0.96 | 0.09 | −0.96 | 0.06 |
| 7 | −1.69 | 0.18 | −1.00 | 0.08 | −1.11 | 0.06 |
| 8 | −1.84 | 0.20 | −1.04 | 0.08 | −1.04 | 0.05 |
| 9 | −2.40 | 0.29 | −1.05 | 0.09 | −0.89 | 0.06 |
| 10 | −1.88 | 0.20 | −0.99 | 0.08 | −0.976 | 0.05 |
| | | | | | | |
| Average | −2.04 | 0.23 | −1.00 | 0.09 | −1.01 | 0.06 |
| Std. err. | 0.26 | | 0.06 | | 0.07 | |

$\widehat{\beta}_{UT}$ is fixed to −20.

**Table 3**
Real data results.

| | RL | RL with LS |
|---|---|---|
| $\widehat{\beta}_{TT}$ | −2.45 | −3.20 |
| Std. err. | 0.06 | 0.06 |
| t-test | −40.83 | −52.59 |
| $\widehat{\beta}_{LT}$ | −1.01 | −1.11 |
| Std. err. | 0.03 | 0.03 |
| t-test | −33.67 | −31.90 |
| $\widehat{\beta}_{LC}$ | −0.40 | −0.32 |
| Std. err. | 0.01 | 0.01 |
| t-test | −40.00 | −21.98 |
| $\widehat{\beta}_{LS}$ | | −0.23 |
| Std. err. | | 0.02 |
| t-test | | −13.20 |
| $LL(\widehat{\beta})$ | 7481.5 | 7213.2 |

$\widehat{\beta}_{UT}$ is fixed to −20.

for the path sampling is used. Frejinger et al. (2009) argue that the resulting parameter estimates are not consistent since the values can vary significantly with the definition of the choice sets. The classic approach has been used for estimating and applying various kinds of route choice models, for example, PSL and c-logit, nested logit (Vovsha and Bekhor, 1998) and mixed logit (Bekhor et al., 2002; Frejinger and Bierlaire, 2007).

The second approach, henceforth referred to as the sampling approach is proposed by Frejinger et al. (2009). It is also path based and choice sets are sampled with some path generation algorithm. However, unlike the classic approach the choice sets are assumed to include all feasible paths in the network and path utilities are corrected for the sampling protocol. The resulting parameter estimates are consistent. The sampling approach has been used for estimating logit and PSL models. Recent results show that also the family of multivariate extreme value models can be estimated on samples of alternatives (Guevara and Ben-Akiva, 2013). The estimated models can be used for prediction by Metropolis–Hastings sampling of paths as proposed by Flotterod and Bierlaire (2013).

The third approach is the link-based one proposed here. The RL model can be consistently estimated and efficiently used for prediction which makes it, in theory, superior to the other models. It is still of interest to empirically compare the RL model and the sampling approach since they can both be consistently estimated and are based on the same assumption of unrestricted choice set. We now turn our attention to discussing how these two approaches can be compared and we start with the estimation results.

If paths with loops are considered by the path sampling algorithm, the path based logit and the RL models are identical. In contrast, if PSL is compared to RL with LS attribute, the parameter estimates may be different because of the correlation correction terms. Since both PS and LS are proxies for correlation it is however unclear which one is the best. The forecasting capabilities would then need to be compared which we discuss in the following.

In order to compare the predictions of the two modeling approaches a real data set should be used. The models can then be estimated based on a subset of the data and applied to a holdout sample. A natural candidate as a performance measure would be the resulting predicted log-likelihood (PLL) values. Computing the PLL for the RL model is straightforward but for the sampling approach it is more involved. Samples of paths need to be used also for the prediction since it is not feasible to

enumerate all paths, even if only simple paths are considered.[6] Note, however, that unlike for estimation, it does not make sense to include the observed paths with probability one in the sample. This implies that an observation can have a predicted probability of zero which is of course problematic. This case actually has a high probability of occurring because the exact observed paths are often not found by path generation algorithms (see coverage results reported e.g. in Ramming, 2002). We conclude that it is not possible to use the state-of-the-art to perform a meaningful forecasting comparison of PSL and RL with LS models. Such a comparison is therefore out of the scope of this paper but it remains an interesting topic for future research.

## 8. Concluding remarks

This paper presents a model for the choice of path in a network with no constraints on the choice set. We start from a dynamic specification of link choices and show that it is equivalent to a static model of the multinomial logit form. We formalize a link size attribute correcting utilities for correlated paths, similar to a path size attribute. The model is estimated by maximum likelihood on simulated and real data in a network with more 3000 nodes and 7000 links. Several instances of simulated data are generated with a chosen model and the resulting parameter estimates are consistent. Moreover, the parameter estimates based on the GPS data have reasonable signs and magnitudes and small standard errors.

There is a large literature on the estimation of the value of travel time, i.e. the willingness-to-pay to reduce travel time.[7] Such estimates are routinely used around the world to evaluate traffic infrastructure investments and thus they guide the use of a noticeable share of GDP in most countries. The variation in the trade-off between travel time and cost is limited in most revealed preference data and therefore most empirical studies of the value of travel time have resorted to stated choices collected in surveys. This situation could change as very rich GPS data sources emerge that allow us to measure the travel times in large networks as well as observe many actual route choices. The framework presented in this paper offers a way to model these choices econometrically.

Future work should be dedicated to explore more ways than the link size attribute to model correlated utilities. One possible extension is a nested approach that allows for different scale parameters, another is to use random parameters.

Baillon and Cominetti (2008) show the existence of a unique equilibrium for a model of the type presented here and provide some numerical experiments. It is an interesting topic for future research to use the recursive logit model in dynamic traffic assignment when their Markovian traffic equilibrium does not apply.

The application in this paper concerns a static network, the model is however valid also for a dynamic network setting as long as the link attributes are deterministic. Ramos et al. (2012) present estimation results for a dynamic model where a state is defined as a link and time pair. They use a GPS data set collected in the The Netherlands where, in addition to path choices, the network link travel times are known at one minute time intervals during the whole data collection period.

## Acknowledgements

## References

Aguirregabiria, V., Mira, P., 2002. Swapping the nested fixed point algorithm: a class of estimators for discrete markov decision models. Econometrica 70 (4), 1519–1543.
Aguirregabiria, V., Mira, P., 2010. Dynamic discrete choice structural models: a survey. Journal of Econometrics 156 (1), 38–67.
Akamatsu, T., 1996. Cyclic flows, markov process and stochastic traffic assignment. Transportation Research Part B 30 (5), 369–386.
Baillon, J.B., Cominetti, R., 2008. Markovian traffic equilibrium. Mathematical Programming 111 (1–2), 33–56.
Beckmann, M., McGuire, C., Winston, C., 1956. Studies in the Economics of Transportation. Yale University Press, New Haven.
Bekhor, S., Ben-Akiva, M., Ramming, M., 2002. Adaptation of logit kernel to route choice situation. Transportation Research Record 1805, 78–85.
Bell, M., 1995. Alternatives to Dial's logit assignment algorithm. Transportation Research Part B 29 (4), 287–295.
Bellman, R., 1957. Dynamic Programming. Princeton University Press, Princeton.
Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short-term travel decisions. In: Hall, R. (Ed.), Handbook of Transportation Science. Kluwer, pp. 5–34.
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 13th International Symposium on Transportation and Traffic Theory. Lyon, France.
DeSerpa, A., 1971. A theory of the economics of time. The Economic Journal 81 (324), 828–846.
Dial, R., 1971. A probabilistic multipath traffic assignment algorithm which obviates path enumeration. Transportation Research 5 (2), 83–111.
Dijkstra, E., 1959. A note on two problems in connexion with graphs. Numerische Mathematik 1, 269–271.
Flotterod, G., Bierlaire, M., 2013. Metropolis-hastings sampling of paths. Transportation Research Part B 48 (1), 53–66.
Fosgerau, M., 2006. Investigating the distribution of travel time savings. Transportation Research Part B 40 (8), 688–707.
Fosgerau, M., McFadden, M., Bierlaire, M., 2013. Choice probability generating functions. Journal of Choice Modelling 8 (1), 1–18.

---

[6] There is an exponential number of simple paths for a given origin–destination pair and enumerating these is a NP-complete problem. Even counting these paths is a #P-complete problem which is at least as difficult as a NP-complete problem (Valiant, 1979).

[7] The micro-economic formulation of the theory of the value of travel time was fundamentally formulated by Beckmann et al. (1956), Johnson (1966), Oort (1969) and DeSerpa (1971). The estimation of the WTP for travel time is reviewed in Hensher (2001) and Gunn (2000). See Fosgerau (2006) for warnings about estimating a mean WTP from discrete data.

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., 2009. Sampling of alternatives for route choice modeling. Transportation Research Part B 43 (10), 984–994.

Guevara, C.A., Ben-Akiva, M.E., 2013. Sampling of alternatives in multivariate extreme value (MEV) models. Transportation Research Part B 48 (1), 31–52.

Gunn, H., 2000. An introduction to the valuation of travel-time savings and losses. In: Hensher, D., Button, K. (Eds.), Handbook of Transport Modelling. Elsevier Science Ltd. (chapter 26).

Hensher, D., 2001. Measurement of the valuation of travel time savings. Journal of Transport Economics and Policy 35 (1), 71–98.

Johnson, B., 1966. Travel time and the price of leisure. Economic Inquiry 4 (2), 135–145.

McFadden, D., 1976. The mathematical theory of demand models. In: Stopher, P., Meyburg, A. (Eds.), Behavioral Travel Demand Models. Lexington Books, pp. 305–314.

McFadden, D., 1978. Modelling the choice of residential location. In: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (Eds.), Spatial Interaction Theory and Residential Location. North-Holland, Amsterdam, pp. 75–96.

Melo, E., 2012. A representative consumer theorem for discrete choice models in networked markets. Economics Letters 117 (3), 862–865.

Oort, C., 1969. The evaluation of travelling time. Journal of Transport Economics and Policy 3 (3), 279–286.

Ramming, M., 2002. Network Knowledge and Route Choice. Ph.D. thesis. Massachusetts Institute of Technology.

Ramos, G., Frejinger, E., Daamen, W., Hoogendoorn, S., 2012. Route choice model estimation in a dynamic network based on GPS data. In: Presented at the 1st European Symposium on Quantitive Methods in Transportation Systems.

Rust, J., 1987. Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. Econometrica 55 (5), 999–1033.

Saad, Y., vander Vorst, H., 2000. Iterative solution of linear systems in the 20th century. Journal of Computational and Applied Mathematics 123 (1–2), 1–33.

Valiant, L.G., 1979. The complexity of enumeration and reliability problems. SIAM Journal on Computing 8 (3), 410–421.

Vovsha, P., Bekhor, S., 1998. Link-nested logit model of route choice overcoming route overlapping problem. Transportation Research Record 1645, 133–142.