CrossMark

Modelling travellers' route choice behaviours with the concept of equivalent impedance

Jun Li[1] · Xinjun Lai[2] iD

Abstract A novel concept is presented to capture route choice behaviours and to account for the correlation of routes in the logit model. The issue that route choice models are easily affected when irrelevant alternatives are included in the choice set is tackled. The concept of equivalent impedance is presented to simplify and aggregate a set of links based on the idea that people tend to remember and process road network information at an abstract level. Then, the equivalent impedance is utilized to derive a correction term for the utility of a multinomial logit model in which the advantages of a closed-form structure and easy computation remain unchanged. The results from the numerical examples suggest that the proposed model obtains reasonable results and provides more stable predictions than comparable models when the composition of the choice set changes. An application in a real urban network with GPS data is presented, and estimation results suggest that the new model is practical due to its robustness.

Keywords Logit route choice · Equivalent impedance · Simplified network · Utility correction · Choice set

Introduction

A route choice model captures travellers' behavioural choices when facing different alternative routes between an origin–destination (OD) pair. The stochastic route choice model provides the probability of each route being chosen; thus, future traffic demand can

✉ Xinjun Lai
  xinjun.lai@gdut.edu.cn

  Jun Li
  stslijun@mail.sysu.edu.cn

[1] School of Engineering, Sun Yat-sen University, Guangzhou 510275, China

[2] School of Electro-Mechanical Engineering, Guangdong University of Technology, Guangzhou 510275, China

be forecasted for transportation planning and management. Due to the closed-form expression and high computational efficiency, the logit model is one of the most commonly used models in route choice analysis.

The multinomial logit (MNL) model is derived based on the assumption that the errors of alternatives are independently and identically extreme value distributed, which leads to a primary drawback: the model cannot interpret the correlation degree of routes due to the independence of the irrelevant alternatives property (IIA), referred to as the overlapping problem, and thus it leads to enlarged probabilities for correlated routes. Several solutions have been proposed to overcome the overlapping problem. The nested logit (NL), cross nested logit (CNL), and paired combinatorial logit (PCL) models interpret the correlation relation explicitly (Chu 1989; Bekhor and Prashker 2001; Bierlaire 2006) in which the correlated routes are classified into the same nest. However, the large set of nesting parameters to be estimated creates a burden for large transportation networks. The mixed logit model incorporates the normal distribution and employs the covariance matrix to interpret the correlation (Bekhor et al. 2007; McFadden and Train 2000; Walker et al. 2001) but the non-closed form requires simulation-based estimation and computation, which are even more time-consuming than the nested logit models. Therefore, the approximation method that adds one correction term to the utility of the MNL model appears to be the most popular in application because of the favourable trade-off between rapid computation and accuracy in prediction. The correlation of routes is considered a disutility to decrease the attraction of overlapped routes and the merits of the MNL model, such as closed-form and efficiency, remain unchanged. We introduce some popular models of this type. The C-logit model (Cascetta et al. 1996) introduced a commonality factor (CF) that is proportional to the correlated parts. The path size logit (PSL) model was proposed with a similar concept (Ben-Akiva and Bierlaire 1999; Frejinger and Bierlaire 2007). The correction term, called the path size (PS), is derived from a property of the Gumbel distribution to aggregate alternatives. Bovy et al. (2008) proposed an alternative method to derive the PS term and named it the path size correction (PSC). These models have been integrated into transportation software, such as Paramics[1] and MITSIM.[2]

The MNL models with utility correction are widely used due to their simple form and efficient computation; however, these models appear to be sensitive to the composition of the route choice set. In most cases, researchers and practitioners can only obtain the chosen route of each trip and must generate the non-chosen alternatives to form the choice set of the model; the choice set that actually considered by each traveller is not always available. Moreover, calculating route choice probabilities also requires the definition of a choice set. Algorithms have been proposed to generate and approximate the choice set considered by travellers but results suggest that it is difficult to include all of the actually chosen routes in the generated set (Ramming 2002; Frejinger and Bierlaire 2007), which means that it is difficult to construct a "perfect" choice set in application. This raises the concern of whether the composition of the choice set affects the performance of route choice models. Moreover, if irrelevant, unreasonably long, or time-consuming routes are included in the choice set, how are the choice probabilities of the most attractive routes affected? Prato and Bekhor analysed this issue (2007) and revealed that the most common route choice models are affected by the composition of choice set and that non-nested logit models yield more robust parameter estimates. Bovy et al. (2008) proposed the PSC-logit model and analysed this choice set issue in non-nested models in which the problem existed. Bliemer and Bovy

---

(2008) tested the same experiment on nested models and found that most route choice models do not appear to have sufficient robustness of choice prediction at the level of individual routes. This is an issue that has been overlooked for years and it appears that few studies have provided solutions. In application, the generation of large choice sets is usually required (Bekhor et al. 2008) and irrelevant routes might be inevitable, indicating the need for a more robust and flexible model to account for this issue. In fact, we mostly focus on the most attractive routes and thus the probabilities of choosing the most attractive routes should ideally be unaffected when unattractive routes are unexpectedly included in the choice set.

We propose a model that is more robust to the composition of choice set and we introduce a novel idea to capture the correlation of routes to better model route choice behaviours. We notice that in route choice analysis there is a trend to simplify the road network to explore travel behaviours. Prato (2009) noted that the degree of complication in transportation analysis increases exponentially with the complexity of the road network and additional study of network simplification approaches might provide more efficiency and accuracy in engineering applications and research. Because travellers do not remember and process the road network in their brain in a link-by-link style and instead tend to process the information in a more abstract style, e.g., using landmarks or major roads, the transportation models should be adjusted accordingly. The concept of a subnetwork was first used in traffic assignment to reduce the computational burden (Xie et al. 2010; Boyles 2012; Barton and Hearn 1979). Frejinger and Bierlaire (2007) used the subnetwork concept in route choice behaviour analysis to simplify the modelling process. The analyst can define a sub-network by choosing major roads from the network hierarchy or by inter- viewing travellers to identify the most frequently used trips. The same idea is adopted by Lai et al. (2016), named the subpath, and employed in a CNL model rather than a mixed logit model to decrease the computational burden. The subpath idea can significantly decrease the number of nests so that the CNL model can be more flexible. Kazagli et al. (2016) propose the mental representation items to interpret these behaviours and describe usages in CNL and recursive logit (Fosgerau et al. 2013). Li et al. (2015) uses the equivalent impedance for traffic assignment. It is practical and rational to build route choice models considering the fact that people may consider utility or impedance of links as a whole or an aggregation. In discrete choice analysis, the Expected Maximum Per- ceived Utility (EMPU), or the inclusive utility, measures individual's expected utility associated with a choice situation (Cascetta et al. 2009). It would be a suitable method to represent travellers' perceived impedance of routes, and intuitively would be a possible means for road network simplification; however, it seems that few attempts succeeded due to the high correlation of routes in the road network. In this regard, we investigate whether a route choice model could be derived based on the concept of network simplification, even though not every network or not the whole network could be simplified. We focus on the MNL model with utility correction because it is the most favourable in application due to its simple form. Therefore, we concentrate on derivation of the correction term. The main contributions to the literature are as follows: (1) we present the concept of the equivalent impedance of a set of links under the logit assumption by which an MNL-based route choice model and the correction term are derived. (2) The proposed model exhibits robustness to the composition of choice set, which has been an issue that has been over- looked for years.

The rest of this paper is organized as follows: "Methodology" section describes the methodology, in which two calculation rules are proposed to merge links in series or in parallel for simplification, based on which the correction term is derived for a route pair.

The routes in the choice set are treated in pairs and the overall correction is determined. Additionally, the variance heterogeneity of the perception error is also discussed. "Numerical example: when a relevant route becomes irrelevant" section provides a three-alternative case, a modified "red/blue bus" network where one relevant route gradually becomes irrelevant, to test the robustness of the new model and compare it with others. "Investigating the choice set composition effect: synthetic data" section presents a grid network and the experiment from Bliemer and Bovy (2008) is used to further investigate the effect of choice set, where estimation and forecasting with synthetic observations are provided. "Application in a real road network" section presents an application in an actual network with GPS data, where estimation and cross-validation are performed to test the practicality of the new model. Conclusions are presented in "Conclusions" section.

Methodology

Consider C_{rs} as the route choice set that includes all the routes from origin r to destination s; a correction term Δ_i is added to the utility function to decrease the attractiveness of the correlated route, which is described by Eq. (1)

$$U_i = V_i + \varepsilon_i = -(\beta x_i + \Delta_i) + \varepsilon_i, i \in C_{rs}, \tag{1}$$

where U_i is the utility of route i, x_i is a vector of attributes, β is the corresponding parameter vector to be estimated, $\beta x_i + \Delta_i$ can be considered as the impedance of route i, namely I_i, $-(\beta x_i + \Delta_i)$ is the deterministic utility, namely V_i, ε_i is the error term of the utility, and all the error terms of the alternative routes in the choice set C_{rs} are Gumbel distributed and independently identically distributed (IID). According to the properties of the EV distribution, the probability that travellers choose route i from choice set C_{rs} can be described by Eq. (2), which is similar to the MNL model with an extra correction term Δ_i,

$$\Pr(i) = \frac{\exp[-\theta_{rs}(\beta x_i + \Delta_i)]}{\sum\limits_{j \in C_{rs}} \exp[-\theta_{rs}(\beta x_j + \Delta_j)]}, \tag{2}$$

where θ_{rs} is the scale parameter of the OD pair rs with a non-negative value and the variance σ_i^2 of $\varepsilon_i, \forall i \in C_{rs}$ is $\sigma_i^2 = \pi^2/6\theta_{rs}^2$.

The derivation of Δ_i in this paper is based on the concept of network simplification. It would be desirable in network analysis if a portion of the road network can be replaced by a link, and then the network could be simplified to some extent. In particular, from the perspective of travellers, the replaced road network and the link should have equal perceived impedances, we term it as the *equivalent impedance*, therefore the travellers would then consider them as equal. If the equivalent impedance of a network can be explicitly calculated, then the difference in the impedances between the network and routes can be compared and the impact of the correlated portion of routes can be determined. We would elaborate the concept of the equivalent impedance in the next section.

Equivalent impedance

The concept of *equivalent impedance* for a set of links is used in the proposed model to derive the correction term in Eq. (2). Two calculation rules are proposed to consider links in series or in parallel.

Links in series

Consider a network in which the links are all in series, as shown in Fig. 1. The deterministic utility of link a is $V_a = -\beta x_a$. It is assumed that the attributes are all link-additive, so that the deterministic utility of the network from r to s is

$$V_{rs} = -\beta \sum_{a \in \Gamma} x_a, \tag{3}$$

where Γ is the set of links. We define $I_{rs} = -V_{rs} = \beta \sum_{a \in \Gamma} x_a$ as the equivalent impedance of the network, which is the impedance of the only route in the network.

Links in parallel

Consider a network that consists of only parallel links connecting r and s (solid lines in Fig. 2). We would like to measure travellers' assessment of the links in this network, and we consider the Expected Maximum Perceived Utility (EMPU), also known as the inclusive utility (Cascetta et al. 2009), is suitable to describe the concept of the equivalent impedance in this case. Therefore, when the underlying assumption on the perception errors of travellers are IID Gumbel, we have

$$I_{rs} = -V_{rs} = -\ln\left[\sum_{a \in \Gamma} \exp(-\theta_{rs}\beta x_a)\right] \Big/ \theta_{rs}. \tag{4}$$

Notice that the formulation of equivalent impedance is not standalone, therefore, we utilize an alternative way to derive the equivalent impedance. An interesting result is discovered that the proposed derivation leads to the same expression as Eq. (4). Besides, the idea of the proposed method is straightforward, and could be easily extended to other cases where the underlying distribution for the error terms is not Gumbel, e.g. it could be Weibull or normal. We elaborate the derivation as follows. Consider a virtual link v connecting r and s with utility $U_v = -\beta x_v + \varepsilon_v$ (dashed line in Fig. 2), such that the probability that travellers choose the virtual link is identical to the probability that travellers choose the original network.

According to this assumption, the error terms ε_v and $\varepsilon_a, \forall a \in \Gamma$, are IID Gumbel distributed; thus, the following holds:

$$\frac{\exp(-\theta_{rs}\beta x_v)}{\exp(-\theta_{rs}\beta x_v) + \sum\limits_{a \in \Gamma} \exp(-\theta_{rs}\beta x_a)} = \frac{\sum\limits_{a \in \Gamma} \exp(-\theta_{rs}\beta x_a)}{\exp(-\theta_{rs}\beta x_v) + \sum\limits_{a \in \Gamma} \exp(-\theta_{rs}\beta x_a)}. \tag{5}$$

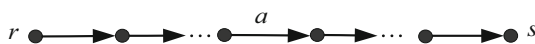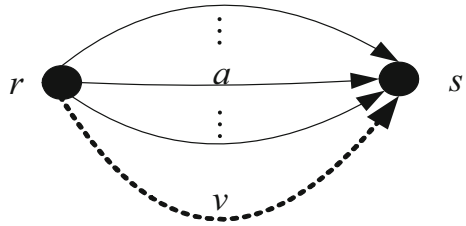By solving Eq. (5), the deterministic part of the utility of link v is

Fig. 1 Network with links in series

Fig. 2 Network with links in parallel

$$V_{rs} = -\beta x_v = \frac{\ln\left[\sum\limits_{a\in\Gamma}\exp(-\theta_{rs}\beta x_a)\right]}{\theta_{rs}}. \tag{6}$$

Because the probability that travellers choose link v is the same as the probability that travellers choose the original network, it is reasonable to use its deterministic utility to represent the *equivalent impedance* of the network, namely

$$I_{rs} = -V_{rs} = -\ln\left[\sum_{a\in\Gamma}\exp(-\theta_{rs}\beta x_a)\right]\Big/\theta_{rs}. \tag{7}$$

Two properties naturally result:

(1) *Monotonicity with respect to link set size* This property implies that any additional parallel link to the network leaves the traveller no worse than before the addition. For example,

$$I_{rs}(a,b) \geq I_{rs}(a,b,c), \tag{8}$$

where $I_{rs}(a,b)$ is the equivalent impedance of the network with links a and b in parallel and $I_{rs}(a,b,c)$ has links a, b and c in parallel.

(2) *Monotonicity with respect to impedance* This property implies that the equivalent impedance of the road network does not decrease if the impedance of any of parallel link increases. For example,

$$\frac{\partial I_{rs}(a,b,c)}{\partial I_k} \geq 0, \forall k \in \{a,b,c\}. \tag{9}$$

These two properties are in accordance with travellers' perceptions and behaviours. Regarding the first property, increasing the number of alternatives in a road network implies more flexibility. Thus, travellers would perceive the road network as less negative and the equivalent impedance will be lower. Regarding the second property, if one route becomes congested, travellers would perceive the road network as being worse, increasing its equivalent impedance.

Note that not all networks can be simplified using the proposed rules but most can be simplified to some degree, which is desirable for derivation of the correction term.

Different distribution assumptions

The sections above present the equivalent impedance in which the underlying model is logit and the error terms are assumed to have a Gumbel distribution. Other models with

different distribution assumptions, such as probit with normal distribution, weibit with Weibull distribution (Yao and Chen 2014), and other distributions (Li 2011; Kitthamkesorn and Chen 2014; del Castillo 2016), can also incorporate this approach when network simplification is needed, especially when the links are in parallel. In particular, some models are inevitably under the assumption of independent distribution, in which the IIA property holds (Kitthamkesorn and Chen 2013), so a customized correction term for overlapping can be derived using the following proposed approach with their own distributions to obtain consistency in model distribution assumptions. If the underlying model changes, e.g., the weibit model with the Weibull distribution, then Eq. (6) would change accordingly.

Even though the derived Eq. (7) has the same expression as the EMPU, however, there are differences between the EMPU and the equivalent impedance. As for the EMPU, it is proposed to measure a perceived maximum value of a set of routes, and thus the motivation is to assess a choice context. On the other hand, as for the equivalent impedance, the motivation is to measure the impedance of a set of links, and thus the goal is to simplify a network. Therefore it could involve a choice context (where the links are in parallel), but it could also *not* involve a choice context (where the links are in series). In the latter case, the EMPU is not applicable. As for the case in "Links in parallel" section, where we employed a choice assumption for the derivation of Eqs. (5)–(7), we consider the EMPU and Eq. (7) are conceptually the same.[3] The concepts of the equivalent impedance and the EMPU are not equivalent, and other formulations that could well explain the concept of the equivalent impedance should be considered and thus would be desirable.

Utility correction term

Correction of a route

The more routes that route i is correlated with, the larger Δ_i should be, which means that every route that route i correlates with should contribute to the total correlation degree of route i. If route i is completely independent, there is no overlapping issue for i and $\Delta_i = 0$. Inspired by the paired combinatorial logit model (Chu 1989), the corrections are treated in route pairs. Any route j that i correlates contributes a disutility Δ_i^{ij} to route i, where Δ_i^{ij} is the correction to route i from route pair ij. All Δ_i^{ij} terms ($\forall j \in C_{rs}, j \neq i$) should account for the overall correlation degree of route i. Moreover, Δ_i^{ij} is only dependent on route i, j and their correlation. The overall correction can be determined as the summation of all paired corrections. Consider route choice set C_{rs} that consists of N routes; thus, there are $N(N - 1)/2$ distinct route pairs. It is natural to assume that the pair corrections are additive. There are several reasons to make the additive assumption: firstly, the more routes that route i correlates, the larger Δ_i should be; secondly, every route that route i correlates should be added for the total correlation degree of route i, it is very difficult to explicitly consider the correlation degree of a large network but it would be more tractable if routes

---

[3] As pointed out by one anonymous reviewer, while the links are in parallel, the inclusive utility and equivalent impedance, as same concept, both allow network simplification.

are treated in pairs. Therefore, we define the total correction of route i is the summation of all corrections in which i exists, which is

$$\Delta_i = \sum_{j \neq i, j=1}^{N} \Delta_i^{ij}. \tag{10}$$

This summation has the same protocol as the PSL and C-logit models to account for the overall correction. Using the PSL model for an example: all the link size factors of PSL are summed as the overall correction for one route, where the corrections from other routes are actually pre-divided and assigned to each link and the overall correction is computed using a link-by-link addition style (Bovy et al. 2008). The additive assumption is employed in a route-based style in this study: the link corrections are considered by the correlated route j, $\forall j \in C, j \neq i$ and the more routes that use link a, the larger the overall correction is. In particular, for the different pairs, the overlapping parts are different so that the correction terms are due to different parts. Therefore, Eq. (10) can be seen as a systematic "counter" to record how many times link a is used by other routes that does not miss or re-count instances.

After the correction, the utility of route i becomes

$$U_i = -\left(\beta x_i + \sum_{j \neq i, j=1}^{N} \Delta_i^{ij} \right) + \varepsilon_i, \forall i, j \in C_{rs}. \tag{11}$$

Derivation of the correction from route pairs

We analyse routes by pair and without loss of generality, two routes i and j from choice set C_{rs}, as shown in Fig. 3, are studied. Routes i and j share a common link a and the independent parts are denoted as b and c, respectively. For illustrative purpose we define in Fig. 3: (1) the overall network of a route pair: the area that composed by the physical links of route pair i and j, $\forall i, j \in C_{rs}$. (2) The sub-network of a route pair: the non-overlapping part of routes i and j.

We assume that the probability that travellers choose route i between rs is equal to the probability that they choose link b between rg, which is

$$\Pr(b \to a | b \to a, c \to a) = \Pr(b|b, c), \tag{12}$$

where the notation $b \to a$ represents the route that uses link a and then link b, which is based on the hypothesis that the impedance of a route is link-additive. Because the goal is
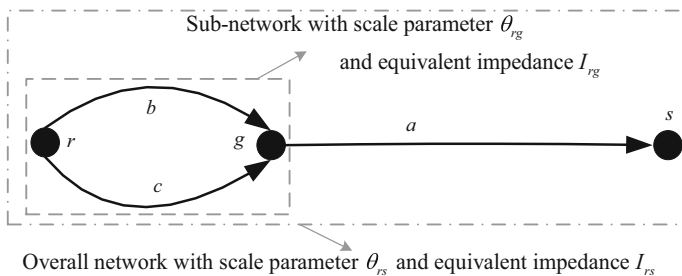

Fig. 3 Two correlated routes. Route i: $b \to a$, route j: $c \to a$

to derive a term to represent the overlapping degree of the routes, it is reasonable to assume that travellers only consider the link-additive attributes when they perceive the degree of overlap in the network. Moreover, the scale parameter θ_{rs} for the overall network of a route pair is different from the scale parameter θ_{rg} of the sub-network.

Note that the assumption of Eq. (11) does not mean that the model ignores the "overlapping" issue of logit models. We emphasize here that only the different portions of the routes are used in the travellers' decision making, which is important for building the new model.

Now consider the sub-network in which the error terms of links b and c are ε_b and ε_c, assumed to be IID Gumbel distributed with the variance $\sigma_{rg}^2 = \pi^2 \big/ 6\theta_{rg}^2$. The utility of links b and c are $U_b = -\beta x_b + \varepsilon_b = -\beta(x_i - x_a) + \varepsilon_b$ and $U_c = -\beta x_c + \varepsilon_c = -\beta(x_j - x_a) + \varepsilon_c$, respectively, thus the right-hand side of Eq. (12) is

$$\Pr(b|b,c) = \frac{\exp\left[-\theta_{rg}\beta(x_i - x_a)\right]}{\exp\left[-\theta_{rg}\beta(x_i - x_a)\right] + \exp\left[-\theta_{rg}\beta(x_j - x_a)\right]}. \tag{13}$$

Denoting the equivalent impedance of the overall network between rs by I_{rs}, then the equivalent impedance of the sub-network between rg is $I_{rg} = I_{rs} - \beta x_a$. According to the calculation rule in "Links in parallel" section, I_{rg} can be calculated using the following equation

$$I_{rg} = I_{rs} - \beta x_a = -\frac{\ln\left\{\exp\left[-\theta_{rg}\beta(x_i - x_a)\right] + \exp\left[-\theta_{rg}\beta(x_j - x_a)\right]\right\}}{\theta_{rg}} \tag{14}$$

and Eq. (14) can be rewritten as

$$\exp\left[-\theta_{rg}(I_{rs} - \beta x_a)\right] = \exp\left[-\theta_{rg}\beta(x_i - x_a)\right] + \exp\left[-\theta_{rg}\beta(x_j - x_a)\right]. \tag{15}$$

Substituting Eq. (15) into the right-hand side of Eq. (13), we have

$$\Pr(b|b,c) = \frac{\exp\left[-\theta_{rg}\beta(x_i - x_a)\right]}{\exp\left[-\theta_{rg}(I_{rs} - \beta x_a)\right]}. \tag{16}$$

The utilities of routes i and j of the overall network are $U_i = -\beta x_i + \varepsilon_i$ and $U_j = -\beta x_j + \varepsilon_j$, respectively. Correction terms are added to modify the utility of each route because routes i and j overlap; therefore, the left-hand side of Eq. (11) becomes

$$\Pr(b \to a|b \to a, c \to a) = \frac{\exp\left[-\theta_{rs}\left(\beta x_i + \Delta_i^{ij}\right)\right]}{\exp\left[-\theta_{rs}\left(\beta x_i + \Delta_i^{ij}\right)\right] + \exp\left[-\theta_{rs}\left(\beta x_j + \Delta_j^{ij}\right)\right]}, \tag{17}$$

where Δ_i^{ij} is the utility correction term to decrease the attraction of route i because of the overlap between routes i and j. Because the equivalent impedance of the overall network between rs is I_{rs}, we have $\exp[-\theta_{rs}I_{rs}] = \exp\left[-\theta_{rs}\left(\beta x_i + \Delta_i^{ij}\right)\right] + \exp\left[-\theta_{rs}\left(\beta x_j + \Delta_j^{ij}\right)\right]$ and therefore Eq. (17) can be rewritten as

$$\Pr(b \to a|b \to a, c \to a) = \frac{\exp\left[-\theta_{rs}\left(\beta x_i + \Delta_i^{ij}\right)\right]}{\exp[-\theta_{rs}I_{rs}]}. \tag{18}$$

Combining Eqs. (12), (16) and (18), we have

$$\frac{\exp\left[-\theta_{rs}\beta x_i - \theta_{rs}\Delta_i^{ij}\right]}{\exp\left[-\theta_{rs}I_{rs}\right]} = \frac{\exp\left[-\theta_{rg}\beta x_i + \theta_{rg}\beta x_a\right]}{\exp\left[-\theta_{rg}I_{rs} + \theta_{rg}\beta x_a\right]}. \tag{19}$$

Equation (19) can be further simplified as

$$\frac{\exp\left[-\theta_{rs}\beta x_i - \theta_{rs}\Delta_i^{ij}\right]}{\exp\left[-\theta_{rs}I_{rs}\right]} = \frac{\exp\left[-\theta_{rg}\beta x_i\right]}{\exp\left[-\theta_{rg}I_{rs}\right]}. \tag{20}$$

Solving Eq. (20), the correction to route i from route pair ij is given by

$$\Delta_i^{ij} = \left(\frac{\theta_{rg}}{\theta_{rs}} - 1\right)(\beta x_i - I_{rs}). \tag{21}$$

Similarly, the correction to route j from correlated link a is

$$\Delta_j^{ij} = \left(\frac{\theta_{rg}}{\theta_{rs}} - 1\right)(\beta x_j - I_{rs}). \tag{22}$$

where $I_{rs} = I_{rg} + \beta x_a = -\frac{\ln\left[\exp\left(-\theta_{rg}\beta x_b\right) + \exp\left(-\theta_{rg}\beta x_c\right)\right]}{\theta_{rg}} + \beta x_a$.

Calculation of scale parameters of sub-networks

The calculation of the correction term requires the scale parameter θ_{rg} from the non-overlapping part of the route pairs; however, this raises a calculation problem: if there are N routes in the choice set, there would be $N(N-1)/2$ pairs of routes, which means there are $N(N-1)/2$ of θ_{rg} to be estimated. An efficient method is to assume that all θ_{rg} are linked to the network topology (Sheffi 1985; Chen et al. 2012), i.e., assuming that the variance of the perception error is proportional to the route length, which means that a shorter route would correspond to a smaller variance. Because the algorithm to find the shortest route is mature and widely used, it would be reasonable to adopt it to interpret the choice context. Therefore, the variance of the non-correlated portion of the shorter route is

$$\sigma_{rg}^2 = \frac{\pi^2}{6\theta_{rg}^2} = \frac{L_{ij}^{\min} - l_{ij}}{L_{ij}^{\min}}\sigma^2, \tag{23}$$

where $\sigma^2 = \pi^2/6\theta_{rs}^2$ is the variance of the perception error between OD, $L_{ij}^{\min}$ is the length of the shorter route in route pair (i,j), and l_{ij} is the length of the overlap between routes i and j. Therefore, the scale parameter θ_{rg} is given by

$$\theta_{rg} = \theta_{rs}\sqrt{\frac{L_{ij}^{\min}}{L_{ij}^{\min} - l_{ij}}}. \tag{24}$$

Setting the scale parameters linked to their impedances decreases the estimation difficulty. If one parameter is fixed, then the others can then be determined, which is desirable in application.

Properties of the route-pair correction term

The proposed route-paired correction term has several properties:

Property 1   If the length of the shortest link in the sub-network, e.g., l_{ij}, goes to zero, the correction term for its corresponding route goes to zero.

$$\lim_{l_{ij}\to 0} \Delta_j^{ij} = \lim_{l_{ij}\to 0}\left(\theta_{rg}/\theta_{rs} - 1\right)(\beta x_i - I_{rs}) = \lim_{l_{ij}\to 0}\left[\sqrt{L_{ij}^{\min}\Big/\left(L_{ij}^{\min} - l_{ij}\right)} - 1\right](\beta x_i - I_{rs}) = 0.$$

(25)

Property 2   The correction terms for two correlated routes in a pair are different and depend on their own impedances.

Property 3   If routes i and j are not correlated, then $\theta_{rg} = \theta_{rs}$ and the correction term Δ_i is zero, resulting in Eq. (2) collapsing into an MNL model.

Property one shows that if unattractive routes are added into the choice set and they have little/no correlation with the relevant routes, the corrections for the attractive routes are small/zero and the probabilities of attractive routes would not be highly affected by the addition of unattractive routes. This property is discussed with an illustrative example in the next section. Property two is useful because the corrections rely not only on their correlation but also on their own impedances (Eqs. (21), (22)), therefore a shorter route with lower impedance has a smaller correction to maintain its attractiveness and vice versa. This is different from the C-Logit, PSL and PSCL models in which the corrections are similar for two routes if they correlate. This property is discussed below with case studies. Property three provides the extreme case of the correction term.

Numerical example: when a relevant route becomes irrelevant

Figure 4 is a revised version of the "red/blue bus" network (Ben-Akiva and Bierlaire 1999), where l_a varies from 0 to 10, the impedance of route k (the lower route) varies from 10 to 20, and route i (the upper route) and route j (the middle route) are always 10. When $l_a = 0$, the three routes have equal lengths and are completely independent; thus, the probability of each route choice should be 1/3 for each alternative. When $l_a = 10$, routes i and j are two separate routes and have the same length of 10 but route k has a length of 20 and is an unattractive route compared with the other two routes. The majority of travellers would choose routes i and j and the probability of choosing each route should be approximately 1/2; few travellers would choose route k.

Five logit-based models and the probit model were selected for comparison with the new model; their formulas and specifications are presented in "Appendix 2". The probit model was set as the reference. The probabilities of choosing route j with varying l_a values are shown in Fig. 5. The curves of the PCL model and the new model almost overlap with

Fig. 4 Experimental network containing three routes. Route i: b, route j: $c \to a$, route k: $d \to a$

$l_b = 10$

$l_c = 10 - l_a$ g s

r

$l_a = 0 \sim 10$

$l_d = 10$
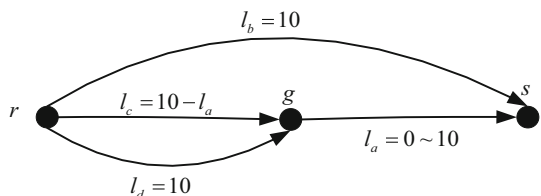
each other (the triangle-labelled and hollow-circle labelled lines, respectively). In Table 1, we provide the root-mean-square error (RMSE) of each route to compare with the results of the probit models and the summation of the RMSE of all routes. The results in Fig. 5 show that the nested models and the new model can provide reasonable results as l_a varies. The three utility-correction models, C-logit, PSL and the PSCL, show their defects as the length of l_a increases when the third route becomes an unattractive choice. The models underestimate the probabilities of choosing route j because it correlates with irrelevant route k. Consequently, the C-logit, PSL and PSCL models overestimate the probabilities of the independent route i. The results in Table 1 show that the model generally has the smallest RMSE compared with the probit, followed by the PCL and the two CNL models. The C-logit, PSL and PSCL models have the largest RMSE. The results suggest that these three utility-correction models have the issue that if the relevant route is "unfortunately" correlated with the irrelevant route, then the choosing probability is affected. In our case, the nested models, the two CNLs and the PCL, have more stable and robust performance than the C-logit, PSL and PSCL models. The new model has very satisfying results regarding the robustness with irrelevant routes, moreover, it retains the simple form of the MNL model with less parameters requiring estimation.

Similar results were discovered by Bovy et al. (2008) that the PSL and PSCL models were sensitive to the composition of the choice set. Following is an explanation for the differences between the traditional utility-correction models, the C-logit, PSL and PSCL models, and the proposed model. If two routes are correlated, the traditional models allocate the same correction values for both. Therefore, if an unattractive route is added to the choice set and correlates with other routes, the probabilities of the attractive routes will be significantly affected. However, the new model calculates the corrections not only based on the correlations but also on their own impedances according to the βx_i term in Eq. (21). Therefore, in the new model, an alternative with a smaller impedance has a smaller correction and vice versa.

As the PCL model and the new model had very similar results in this test, we further compared models in the original "red/blue bus" network, where $l_d = 10 - l_a$ and all routes have the same length. The results are shown in Fig. 6. All models reproduced the trend that when $l_a = 0$, all routes are independent, so the choosing probabilities are all 1/3; when
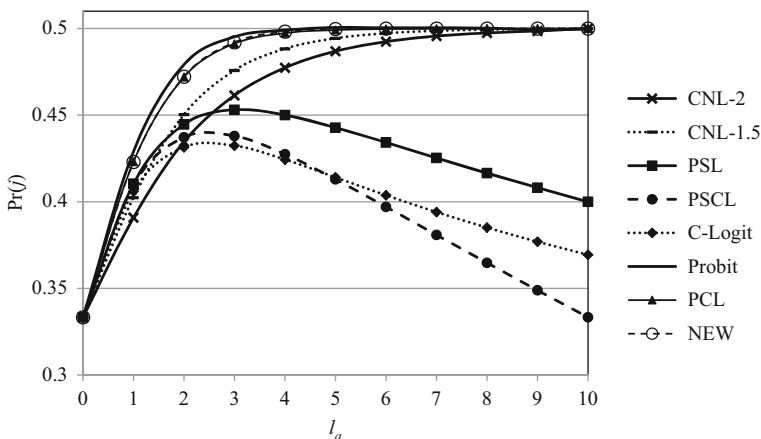

Fig. 5 Probabilities of choosing route j calculated using different models

Table 1 RMSE of the probabilities of choosing routes i, j and k (probit model used as a reference)

RMSE	PSL	PSCL	C-logit	PCL	CNL-1.5	CNL-2	New
Pr (i)	0.601	0.975	0.861	0.0277	0.124	0.213	0.0315
Pr (j)	0.630	0.995	0.882	0.0323	0.144	0.231	0.0304
Pr (k)	0.0815	0.0572	0.0489	0.0236	0.267	0.443	0.0121
Sum	1.31	2.03	1.79	0.0836	0.535	0.887	0.0740

$l_a = 10$, routes j and k are 100% correlated so both have a choosing probability of 1/4. The new model has the most similar curve to the probit and the PCL model shows less satisfying results. This suggests that the new model performs better than the PCL model in terms of resolving the overlapping problem in this network. Therefore, the PCL and the new model perform well when a route in the choice set is irrelevant but the new model outperforms the PCL in modelling correlation of routes.

Investigating the choice set composition effect: synthetic data

We investigate the choice set issue in a grid network and estimate and apply the models using synthetic data generated by the probit model. The experiment is the same as that conducted by Bliemer and Bovy (2008) to study how the composition of choice set affects model performance. The motivation of the experiment is that, in application, we estimate a model with choice set C_e and then apply the model for forecasting with choice set C_f, where irrelevant routes are inevitably included in C_e and C_f. Therefore, we analyse how models perform if: (1) C_e only includes relevant routes and C_f includes irrelevant routes and (2) C_e includes irrelevant routes and C_f only includes relevant routes. The simulated results from the probit model serve as the observations for estimation and for comparison with the prediction results. We refer readers to Bliemer and Bovy (2008) for more information on the design of the experiment.
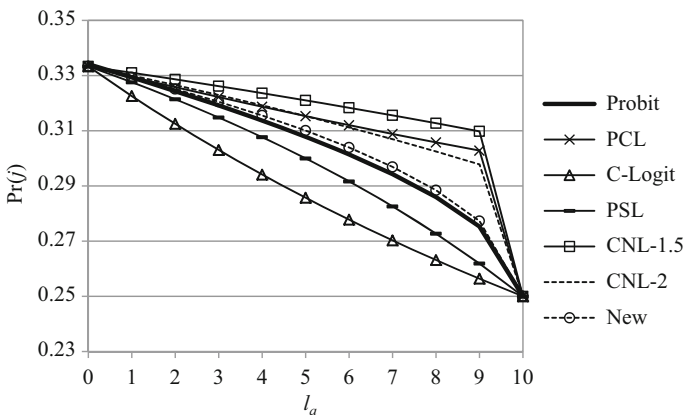

Fig. 6 Comparison of models in the original "*red/blue* bus" network

The grid network shown in Fig. 7 is a single OD pair from origin 1 to destination 16. The length of each link is marked by a double arrow. Four types of models are compared in this case: the probit model (as the "true value"), the traditional MNL models with/without utility modification: the logit, PSL, PSCL and C-logit models, the PCL and CNL models (because they exhibited good performance in "Numerical example: when a relevant route becomes irrelevant" section), and the new model.

Choice set generation

Alternative routes are generated by enumeration and sorted by length. We consider routes 1–5 to be the most relevant routes, routes 6–11 are relevant routes, and routes 12–25 are the irrelevant routes. Three choice sets of different sizes are considered, named C_1, C_2 and C_3, as shown in Table 8 in "Appendix 1.1".

Experiment specification

The experiment was carried out by: (1) Generating synthetic observations with the probit model and C_1 and estimating the parameters of models using these postulated data. (2) Using the estimated models to compute route choice probabilities with choice sets C_1, C_2 and C_3. (3) Comparing the computation results with the probit model. The probabilities computed using the probit model are calculated with a Monte Carlo simulation with 10,000 repetitions. The simulation results from the probit model are assumed to be the synthetic data (postulated observations), which are used for estimation of the models compared. The variance of the probit model is assumed to be the same as that of the logit model to define the scale parameter. The postulated probabilities are shown in Fig. 12 in "Appendix 1.2".

We first estimate the models with choice set C_1 and the corresponding probit data, then compute the probabilities of C_1, C_2 and C_3. The results are shown in Fig. 13 in "Appendix 1.2", where the x-axis is the identification number of the routes in the choice set (the first, second and third columns show C_1, C_2 and C_3) and the y-axis shows the probit probabilities. We then estimate the models with choice set C_3 and compute the probabilities of C_3, C_2 and C_1. The results are shown in Fig. 14 in "Appendix 1.2". Thus, the estimation–

Fig. 7 Study network with a single origin–destination pair from 1 to 16
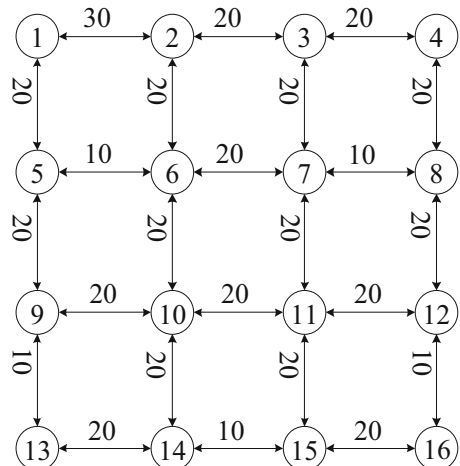
computation exercises were performed twice. Although the data set for estimation and prediction is the same, it does not contradict with the motivation of this experiment. A formal cross-validation with different datasets from actual observations is presented in "Cross-validation" section.

We compare the RMSE of the models against that of the probit, the "true" model, and a lower RMSE indicates better performance. Apart from the purely statistical considerations, correlation among routes has behavioural impacts, therefore, we also estimate the correlation parameters: for the PSL, PSCL, C-Logit and the new models, this parameter is for the correction term; for the PCL, it is for the similarity index; for the CNL, it is the nest scale parameter in which all are assumed to be the same.

Analysis of robustness

The RMSE of models are reported in Table 2: the second and last rows can be interpreted as the "goodness-of-fit" in estimation and the other rows can be interpreted as the computation precision.

Regarding "C_1 (estimated)", the results suggest that the CNL model obtains the best goodness-of-fit, followed by the PSL and C-logit models, the new model, and the PSCL and the PCL models perform the worst. For C_2 and C_3, the new model has average performance among the compared models: the 4th smallest RMSE in C_2 and the 2nd smallest RMSE in C_3.

Regarding "C_3 (estimated)", the proposed model has the 3rd smallest RMSE in goodness-of-fit and appealing prediction results: the second best in C_2 and the best in C_1, indicating its robustness when irrelevant routes are included in the choice set. In both cases, except for the new model, the MNL-structure models show unstable performances as the choice set changes, as the probabilities of the most irrelevant routes are affected by removal or inclusion of irrelevant routes. The PCL model, which has good performance in the three-alternative case, does not perform well in this test. The CNL model with the estimated parameters has satisfying results in this case for both estimation and prediction, which suggests satisfying robustness.

We compare the probabilities of the five shortest routes computed by the CNL and new models and the results are shown in Fig. 8. As the choice set diminishes (Fig. 8a), the probabilities of the shortest routes generally increased and vice versa (Fig. 8b). It appears that the results of the proposed model are more stable when the choice set changes, we provide absolute differences in probabilities. Figures 13 and 14 in "Appendix 1.2" show more detail in the variations as the choice set changes.

Table 2 RMSE compared with the probit model

	Logit	PSL	PSCL	C-Logit	PCL	CNL	New
C_1 (estimated)	0.0683	0.0590	0.137	0.0644	0.114	0.0363	0.0768
C_2	0.0742	0.0358	0.0996	0.0500	0.121	0.0394	0.0709
C_3	0.0801	0.0831	0.0476	0.0561	0.111	0.0389	0.0455
C_1	0.0597	0.0311	0.0516	0.0295	0.0395	0.0108	0.0307
C_2	0.0636	0.174	0.209	0.0997	0.116	0.0268	0.0325
C_3 (estimated)	0.0695	0.0914	0.260	0.0511	0.155	0.0390	0.0293

These findings are similar to those of previous studies (Bovy et al. 2008; Bliemer and Bovy 2008) in which the models with a utility correction were found to be sensitive to the composition of the choice set. We test these models with a new grid network and re-validate their findings. The results also show that the proposed model exhibits robust performance in this test and satisfying results in both estimation and prediction, suggesting that it is suitable for application.

Application in a real road network

The city studied, Guangzhou, is in southern China and contains approximately 10 million inhabitants and 20,000 taxis. Only the central business district, the screenshot shown in Fig. 9, is studied. Observations were collected using GPS devices from urban taxis when they were transporting passengers. The GPS devices were installed by a management company for monitoring purposes, not for navigation; thus, the route choice behaviours were based on the drivers' own judgments and not from navigation suggestions. Information on the network and the data collected are shown in Tables 3 and 4.

Model specification

Four attributes, shown in Table 5, are chosen for the utility functions. *Length* and time are two highly similar and correlated attributes, so one is sufficient for the utility function. However, a precise travel time is difficult to obtain before departure. When drivers decide which route to choose, they generally process the information that they consider to be more stable and distance is relatively more stable than the travel time in the example, so the distance is employed in the utility function rather than the time. The *Artery Road Ratio* is the ratio of artery road in the trip. It is used to test the assumption that travellers prefer to drive on an artery road and a higher ratio is expected to have larger utility. The *Nb. of Signal*- and *Non-signal*-controlled intersections are expected to have significant effects on urban route choice.
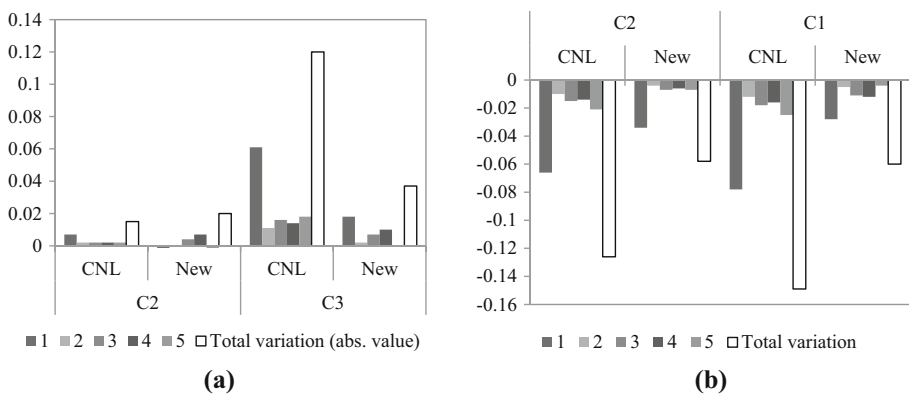

Fig. 8 Variations in probabilities of the five shortest routes when the choice set changes. **a** C_1 is estimated, **b** C_3 is estimated
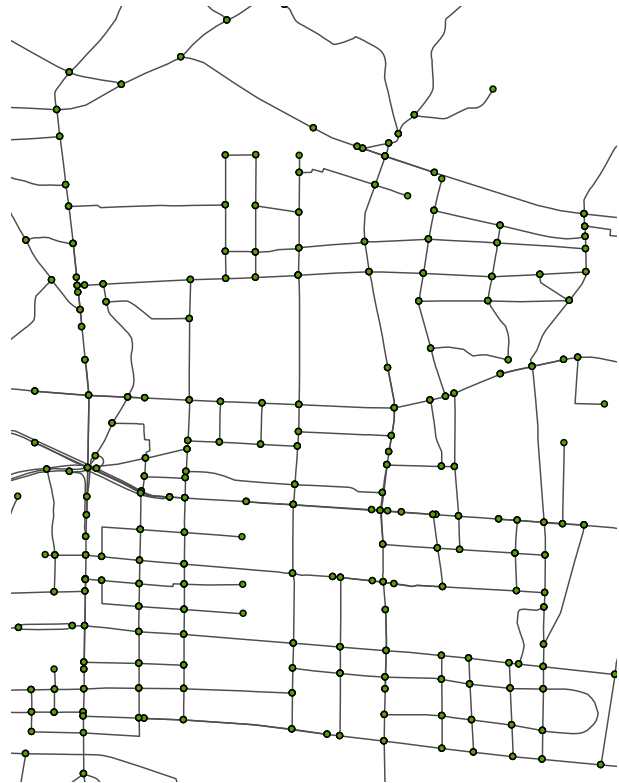
Fig. 9 Map of the region of Guangzhou city studied

Table 3 The network

Nodes	Unidirectional links	Major roads	Arterial streets	Minor streets	Signal-controlled intersections	Non-signal-controlled intersections
208	662	24	34	32	57	208

Table 4 Data collected

Taxis	Observations	OD pairs
2569	7810	1066

The MNL, PSL, PCL, logit kernel (LK) and new models are compared in this case. The LK model is selected because it is difficult to use the probit model for a large road network but the LK model has an error term that consists of two parts: an IID Gumbel distributed part and normally distributed one. It can be interpreted as a mixture of the logit and probit models. The formula and specification of the LK model are in "Appendix 2"; it requires a simulation-based method for estimation because of the non-closed form expression.

Table 5 Statistics for routes corresponding to observations

Attributes	Min	Average	Max
Length (km)	0.447	2.842	8.342
Artery road ratio	0	0.7973	1
Number of non-signal-controlled intersections	3	11.4563	31
Number of signal-controlled intersections	0	3.6571	14

The utility is specified with a linear-in-parameters formulation. Apart from the purely statistical considerations, correlation among routes may have behavioural impacts, so the impact of the Δ and lnPS factors on the choice behaviour should be described by the parameters β_D (for Δ) and β_{PS} (for ln PS) that are estimated from observations (Frejinger and Bierlaire 2007). We only consider length as the impedance to compute Δ. The deterministic utility V_i for the alternative i in the new model is shown in Eq. (26), which is different from Eq. (27) because the to-be-estimated parameter β "absorbs" the negative sign.

$$V_i = \theta_{rs}(\beta_D \times \Delta_i + \beta_L \times Length_i + \beta_{ARR} \times ArteryRoadRatio_i + \beta_{NS} \times NonSignal_i + \beta_S \times Signal_i)$$

(26)

Model estimation

Estimations of parameters are provided in Table 6. The choice sets are generated by link elimination (Azevedo et al. 1993) and the chosen route is added to the choice set if it is not included in the generated choice set. Two choice sets were used for the analysis, one with 30 routes and one with 50 routes, based on the fact that 30 is large enough for a traveller to process and 50 is even larger. The large route choice sets enable investigation of whether the model is too sensitive to those irrelevant routes because the larger the choice set is, the more likely that the choice set includes unrealistic and irrelevant routes.

Most of the estimates have the expected signs, except for the parameter of *Nb. of non-signal intersections*: the models all have positive signs except for the logit and LK models with larger choice sets that have the expected negative sign. A possible explanation is that the larger *Nb. of non-signal intersections* suggests more flexibility in the alternative route and is therefore positive for travellers' utility. Moreover, the *Nb. of signal intersections* has a more negative and dominant impact on travellers' utility than the *Nb. of non-signal intersections*, so the former should be negative but not necessarily the latter.

The new model, the MNL and PSL models share an MNL-like structure, and they have same magnitudes for different estimates. Regarding the new model, the estimates are significantly different from 0 as the t test results are greater than 1.96. Moreover, the estimate for the correction term has the expected negative sign, suggesting a disutility to decrease the attraction of a correlated route. The estimation times of the MNL, PSL and new models were only 1 or 2 sec but the PCL model took approximately 10 min and the LK model took 1–2 h. The new model has the advantage of fast computation due to the MNL structure; this feature is useful when applied to traffic assignment in a large-scale network.

Table 6 Estimation results (30 or 50 alternatives in the choice set)[a]

Coefficient (scaled estimate)	30 routes in the choice set						50 routes in the choice set					
	MNL	PSL	PCL	CNL	LK	New	MNL	PSL	PCL	CNL	LK	New
Artery road ratio	5.71	5.65	6.64	3.28	3.29	5.16	5.81	5.78	7.24	3.37	5.06	5.42
Std. error	0.0975	0.0840	0.0939	0.0255	0.0266	0.0821	0.101	0.0850	0.0980	0.0267	0.0102	0.0827
t-test	58.5	67.3	70.8	128	123	62.9	57.8	68.0	73.9	126	494	63.2
Length	−1.39	−1.36	−3.76	−1.05	−1.04	−1.24	−1.60	−1.51	−3.96	−1.17	−2.14	−1.37
Std. error	0.0617	0.0492	0.0272	0.0153	0.0246	0.0493	0.0603	0.0490	0.0259	0.0153	0.0101	0.0492
t-test	22.5	27.6	138	68.8	42.4	25.1	26.5	30.7	153	76.3	212	27.8
Non-signal-controlled intersections	−0.0811	0.134	0.811	0.0021	−0.806	0.0529	−0.0939	0.106	0.826	0.0048	0.0889	0.0550
Std. error	0.0119	0.0144	0.0212	0.0068	0.0242	0.0125	0.0122	0.0146	0.0206	0.0068	0.0153	0.0126
t-test	6.80	9.31	38.2	0.292	33.3	4.24	7.67	7.22	40.1	0.681	5.80	4.37
Signal-controlled intersections	−0.805	−0.538	−0.644	−0.365	−1.09	−0.557	−0.816	−0.578	−0.420	−0.376	−0.167	−0.560
Std. error	0.0165	0.0172	0.0255	0.0083	0.0525	0.0163	0.0170	0.0173	0.0249	0.0085	0.00881	0.0163
t-test	48.8	31.2	25.3	44.1	20.8	34.1	47.9	33.3	16.9	44.4	18.9	34.4
λ_{LK}, λ_{PCL} or θ_m	–	–	0.399	5.67	−0.158	–	–	–	0.445	5.48	−1.11	–
Std. error			00271	0.0771	0.0162				0.0337	0.0706	0.0602	
t-test			14.7	73.6	9.75				13.2	77.5	18.4	
Δ or ln(PS)	–	0.569	–	–	–	−0.390	–	0.556	–	–	–	−0.333
Std. error		0.0229				0.0153		0.0246				0.0122
t-test		24.8				25.5		22.6				27.4

Table 6 continued

Coefficient (scaled estimate)	30 routes in the choice set						50 routes in the choice set					
	MNL	PSL	PCL	CNL	LK	New	MNL	PSL	PCL	CNL	LK	New
Nb. of estimate parameters	4	5	5	5	5	5	4	5	5	5	5	5
Null log-likelihood	−23,625	−23,625	−23,625	−23,625	−23,625	−23,625	−27,173	−27,173	−27,173	−27,173	−27,173	−27,173
Final log-likelihood	−12,267	−11,971	−14,146	−11,150	−12,085	−11,955	−12,545	−12,297	−14,617	−11,329	−12,442	−12,179
$\bar{\rho}^2$	0.481	0.493	0.401	0.528	0.488	0.494	0.538	0.547	0.462	0.583	0.542	0.552
Estimation time	1 s	1 s	4.45 min	6.00 min	1.18 h	1 s	2 s	2 s	12.8 min	8.98 min	1.65 h	2 s

[a] Number of observations: 6946; LK is simulated with 1000 draws; estimation was run using R2014a Matlab in Windows with an Intel i7 CPU and 8 GB RAM

The goodness-of-fit of models is compared using the adjusted likelihood ratio index $\bar{\rho}^2$. A higher $\bar{\rho}^2$ indicates a better model fit. In both cases, the CNL model has the highest $\bar{\rho}^2$, followed by the new model, then the PSL and MNL. The PCL and LK models have the worst goodness-of-fit, even worse than the MNL model. The weak performance of the LK model may be due to the small number of draws for simulation. The results from two choice set sizes suggest that a larger choice set has a better goodness-of-fit as the $\bar{\rho}^2$ is higher.

The CNL and the new model have the best goodness-of-fit in estimation. Note that only the log-likelihood function of the new model can obtain the global maximum because of the linear assumption in the utility function, therefore the estimation process is fast. Conversely, only a local maximum can be achieved with the CNL model because of its complicated structure, moreover, the estimates are very sensitive to the initial value settings and upper and lower bounds. In our case, in which all the nesting scales are assumed to be the same, the estimation of this single scale requires repeated adjustment of the settings. Estimation of additional scale parameters, e.g., the individual scale for each nest, would not be easy and would likely fail. In this sense, the MNL-structure of the new model has an obvious advantage in application.

Cross-validation

We perform a cross-validation test to further investigate the forecasting abilities of the models. The MNL, PSL, CNL and new models are compared because of their goodness-of-fit and shorter computation times. The observations are divided into two parts: 80% of the OD pairs are randomly selected for estimation and the remaining 20% are for forecasting validation. This was carried out 100 times and the results are presented in two boxplots in Figs. 10 and 11. "CC = 30" and "CC = 50" represent the choice set sizes of 30 and 50. The averages of the log-likelihoods of 100 repetitions are reported in Table 7.

The new model performs the best in this test, as its median and average of log-likelihoods in estimation and forecasting-validation are the largest among the compared models for both choice set sizes. The PSL, CNL and the new model exhibit better performances than the MNL model, suggesting that considering the correlations of routes interprets the route choice behaviours better. Although the CNL model has the best goodness-of-fit in
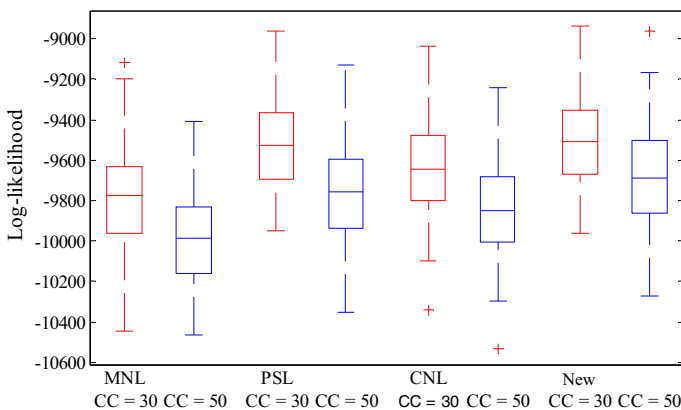

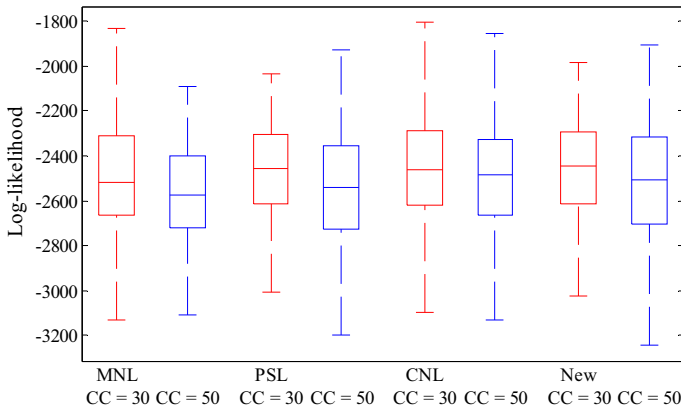Fig. 10 *Boxplot* of estimation log-likelihoods (100 repetitions)

Fig. 11 *Boxplot* of forecasting-validation log-likelihoods (100 repetitions)

Table 7 Average log-likelihood of 100 times of cross-validation

Average log-likelihood	30 routes in the choice set				50 routes in the choice set			
	MNL	PSL	CNL	New	MNL	PSL	CNL	New
Estimation	−9772	−9515	−9628	−9504	−9983	−9773	−9833	−9693
Forecasting	−2497	−2464	−2466	−2457	−2568	−2537	−2506	−2498

estimation with the full data set, it does not outperform the PSL and new models in this test. Summarizing the estimation and cross-validation results in this case study, we conclude that the new model provides reasonable and practical results: it has an obvious advantage in computational efficiency due to its MNL structure, and it exhibits robustness to the composition of choice set and estimation and forecasting results suggest that it is a suitable model for application.

Conclusions

In this paper, the concept of equivalent impedance is proposed to simplify a set of links in a road network, with the motivation that travellers remember, perceive, and process road network information at an abstract and conceptual level. According to this idea, we propose a novel route choice model that aims for greater robustness and practicality in capturing the route choice behaviours of travellers. A correction term is derived based on the proposed concept of equivalent impedance to account for the correlation of routes in an MNL route choice model. A major contribution of this paper is the proposal of a new model that tackles a persistent issue in the route choice literature: models should be robust for the composition of the choice set.

Three tests are presented in this paper: a revised version of the "red/blue bus" network, a grid network with postulated observations, and an actual network with GPS data. Among the models compared, the new model exhibits reasonable and satisfying outcomes. The results from "toy" networks exemplify the model's robustness in providing stable results

when adding irrelevant routes to the choice set, and it appears to be less sensitive to the choice set composition than other models. The results from the actual case suggest that the new model is suitable in application and generally outperforms other models in tests.

The new method explores additional possibilities for simplifying networks: it explicitly calculates the utility for a set of links and may provide some perspective for future research on complex urban networks. An advantage of this model is that it is also useful for a wider range of other domains in discrete choice analysis, e.g., residential choice and online product choice, in which the actual choice set is difficult to obtain and the choice set could be inevitably large, and thus, the robustness of the proposed model would be desirable.

A possible future direction of the proposed method is to relax the assumption in Eq. (23), in which the scales are assumed to be linked rather than estimated from observations to reduce the complexity of large network analysis. This is a common problem encountered by many route choice models, such as the CNL and mixed logit models. A possible solution is to model the route correlation using a more abstract rather than a link-specific manner, such as the sub-path method with the CNL model (Lai et al. 2016) and the subnetwork approach with mixed logit models (Frejinger and Bierlaire 2007). Another possible direction would be to explore the correction terms for other models, such as weibit, in which the independently Weibull distributed assumption is made for its error terms, and thus, the model cannot account for the correlation degree in route choice (Kitthamkesorn and Chen 2013). The proposed method provides a possible solution to derive a correction term that is based on the Weibull distribution. Besides, we consider the formulations for the equivalent impedance are not standalone, and more explorations on this concept would be desirable. In particular, a systematic method to compute the equivalent impedance of a complicated network, where routes are highly correlated, would be valuable.

Appendix 1

Appendix 1.1

See Table 8.

Table 8 The 25 shortest routes generated by enumeration

Route ID	Length	Node sequence							Choice sets		
1	90	1	5	6	7	8	12	16	C_3	C_2	C_1
2	100	1	5	6	7	11	12	16			
3	100	1	5	6	10	14	15	16			
4	100	1	5	6	10	11	12	16			
5	100	1	5	9	13	14	15	16			
6	110	1	5	6	10	11	15	16			

Table 8 continued

| Route ID	Length	Node sequence									Choice sets
7	110	1	5	6	7	11	15	16			
8	110	1	5	9	10	14	15	16			
9	110	1	2	3	7	8	12	16			
10	110	1	2	6	7	8	12	16			
11	110	1	5	9	10	11	12	16			
12	120	1	5	9	10	11	15	16			
13	120	1	2	6	7	11	12	16			
14	120	1	2	3	7	11	12	16			
15	120	1	2	6	10	14	15	16			
16	120	1	2	3	4	8	12	16			
17	120	1	2	6	10	11	12	16			
18	130	1	2	6	10	11	15	16			
19	130	1	2	6	7	11	15	16			
20	130	1	2	3	7	11	15	16			
21	130	1	5	6	10	11	7	8	12	16	
22	130	1	5	6	10	9	13	14	15	16	
23	130	1	5	6	10	14	15	11	12	16	
24	130	1	5	6	2	3	7	8	12	16	
25	130	1	5	9	13	14	15	11	12	16	

Appendix 1.2

See Figs. 12, 13, and 14.
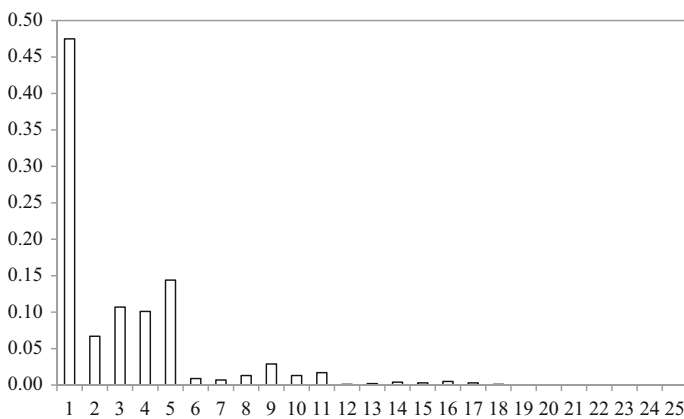
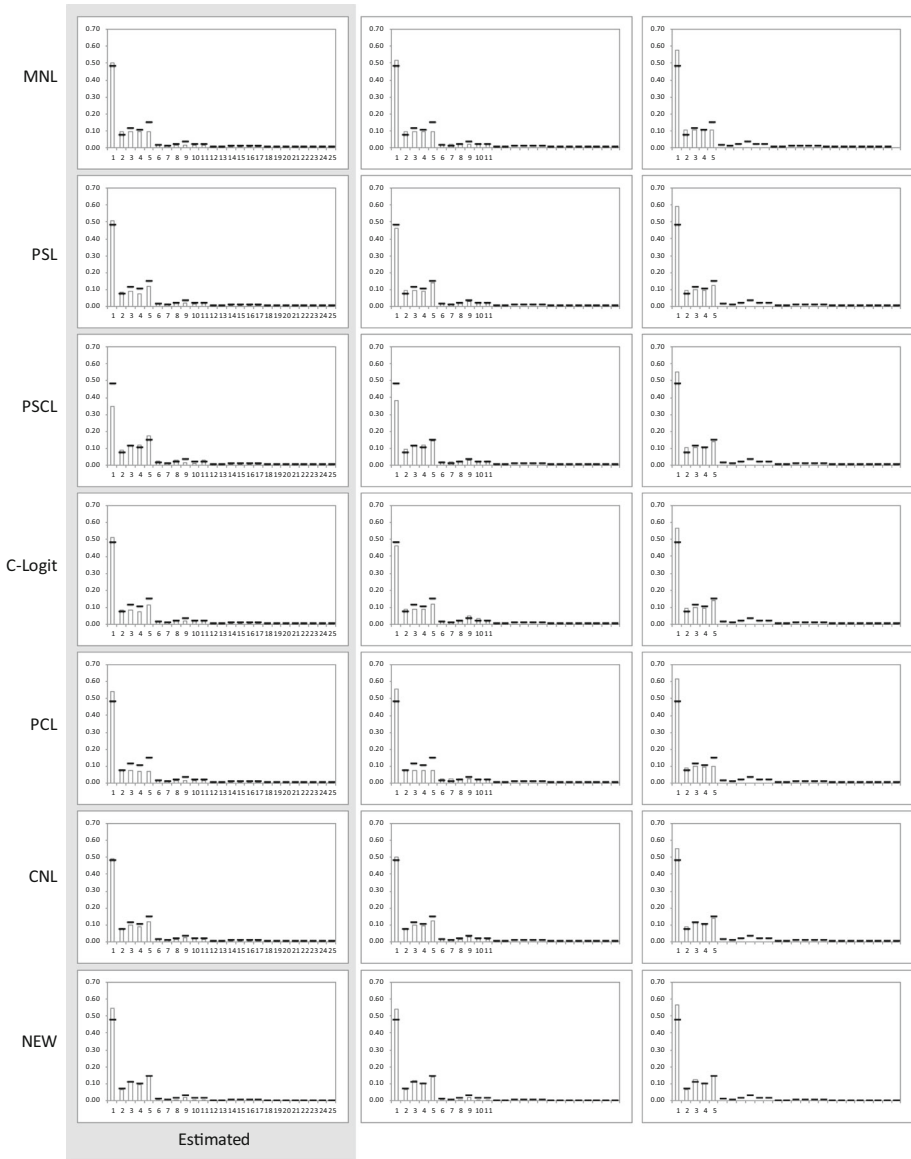**Fig. 12** Simulated probit probabilities of 25 routes in the grid network

Fig. 13 Probabilities of routes compared with the probit model (the *bar* label) as the choice set changes (x-axis); C_1 is estimated
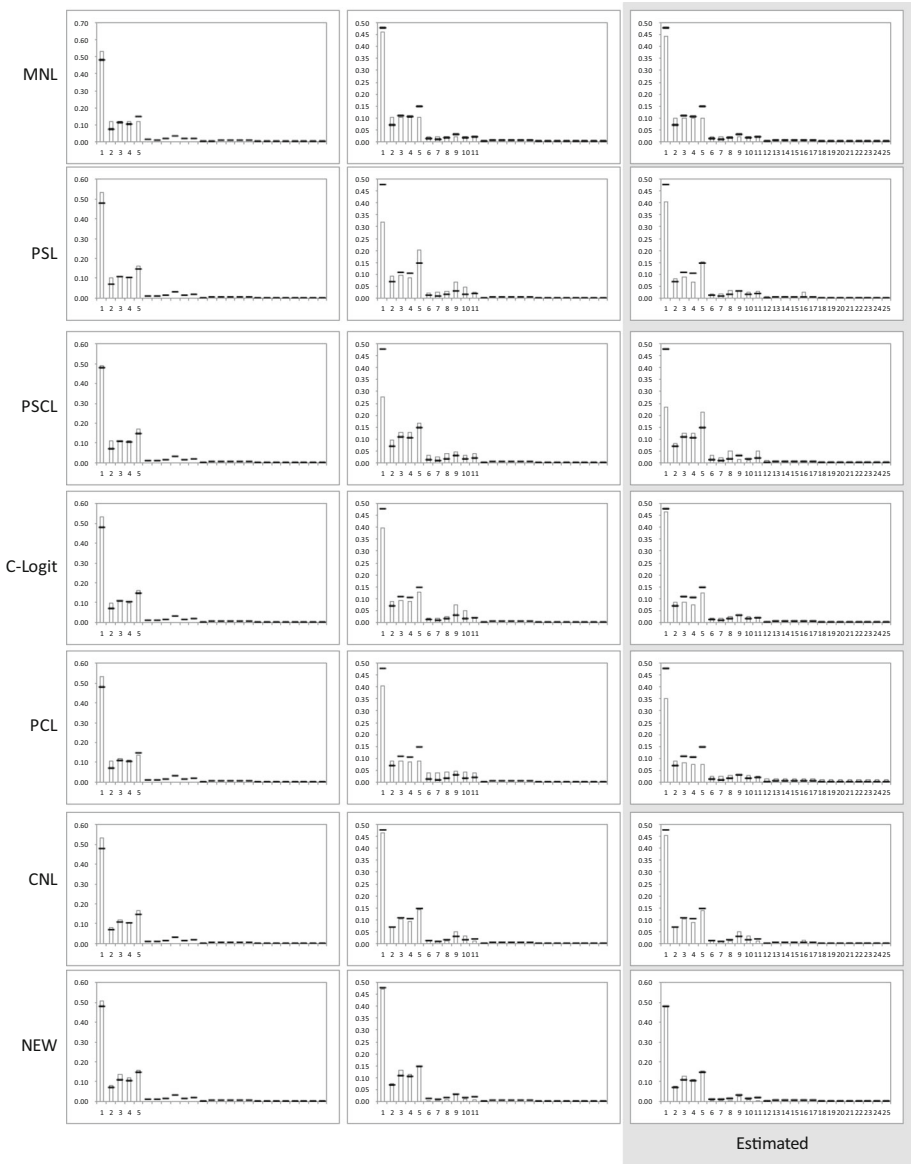
Fig. 14 Probabilities of routes compared with the probit model (the *bar* label) as the choice set changes (x-axis); C_3 is estimated

Appendix 2

(1) The models compared in "Numerical example: when a relevant route becomes irrelevant" section are:

- *Multinomial logit models with utility correction*, including path size logit (PSL), C-logit and path size correction logit (PSCL) models. The probability of choosing route i from choice set C_{rs} is formulated as follows:

$$\Pr(i|C_{rs}) = \frac{\exp[-\theta_{rs}(\beta x_i + \beta_{corr} corr_i)]}{\sum_{j \in C_{rs}} \exp[-\theta_{rs}(\beta x_j + \beta_{corr} corr_j)]}, \qquad (27)$$

where $corr_i$ is the utility-based correction term for route i, β_{corr} is to be estimated, and $corr_i$ from the C-logit model is called the commonality factor and can be expressed as (Cascetta et al. 1996)

$$corr_i = CF_i = \ln \sum_{i \in C_{rs}} \left(l_{ij} / \sqrt{L_i L_j} \right), \qquad (28)$$

where l_{ij} is the length of the overlap between routes i and j and L_i and L_j are the lengths of routes i and j, respectively. The correction terms in the PSL (Ben-Akiva and Bierlaire 1999) and PSCL (Bovy et al. 2008) models, denoted as PS and PSC, respectively, can be expressed as

$$corr_i = \ln(PS_i) = \ln \left(\sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_{rs}} \delta_{aj}} \right) \qquad (29)$$

and

$$corr_i = PSC_i = -\sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i} \ln \frac{1}{\sum_{j \in C_{rs}} \delta_{aj}} \right) \qquad (30)$$

where PS_i is the path size for route i, PSC_i is the path size correction term for route i, Γ_i is the set of links belonging to route i, δ_{aj} is a route-link incidence dummy which is equal to one if route j uses link a and zero otherwise. The estimated $corr_i$ parameters for the PSL and PSCL models are expected to be negative whereas the value for the C-logit model is expected to be positive.
- *The paired combinatorial logit model* (Chu 1989) in which the routes are compared in pairs. The probability of choosing route i is formulated as

$$\Pr(i|C_{rs}) = \sum_{j \in C_{rs}, j \neq i} \Pr(ij) \Pr(i|ij), \qquad (31)$$

where $\Pr(i|ij)$ is the conditional probability of choosing route i given a chosen pair ij:

$$\Pr(i|ij) = \frac{\exp\left(-\frac{\theta_{rs}\beta x_i}{1-\sigma_{ij}}\right)}{\exp\left(-\frac{\theta_{rs}\beta x_i}{1-\sigma_{ij}}\right) + \exp\left(-\frac{\theta_{rs}\beta x_j}{1-\sigma_{ij}}\right)}. \qquad (32)$$

$\Pr(ij)$ is the marginal probability of choosing pair ij:

$$\Pr(ij) = \frac{\left[\exp\left(-\frac{\theta_{rs}\beta x_i}{1-\sigma_{ij}}\right) + \exp\left(-\frac{\theta_{rs}\beta x_j}{1-\sigma_{ij}}\right)\right]^{1-\sigma_{ij}}}{\sum_{k=1}^{N-1}\sum_{l=k+1}^{N}\left[\exp\left(-\frac{\theta_{rs}\beta x_k}{1-\sigma_{kl}}\right) + \exp\left(-\frac{\theta_{rs}\beta x_l}{1-\sigma_{kl}}\right)\right]^{1-\sigma_{kl}}}, \tag{33}$$

where N is the number of routes in C_{rs}. The similarity index σ_{ij} is defined as $\sigma_{ij} = \left(l_{ij}/\sqrt{L_iL_j}\right)^{\gamma}$ based on the topology of the network (Prashker and Bekhor 1998).

- *A link-based CNL model* (Prashker and Bekhor 1998; Vovsha and Bekhor 1998) in which each link can be considered a nest and a route that uses a link belongs to the nest, which can be expressed as:

$$\Pr(i|C_{rs}) = \sum_{m \in \Gamma} \Pr(i|m)\Pr(m), \tag{34}$$

where Γ is the link set, $\Pr(i|m)$ is the conditional probability of choosing route i given the nest m is chosen, i.e., route i uses link m, and $\Pr(m)$ is the marginal probability of choosing nest (link) m. The following are the formulations for the conditional and marginal probabilities:

$$\Pr(i|m) = \frac{\alpha_{im}\exp(-\theta_m\beta x_i/\theta_{rs})}{\sum_{j\in C_m}\alpha_{jm}\exp\left(-\theta_m\beta x_j/\theta_{rs}\right)} \tag{35}$$

$$\text{and } \Pr(m) = \frac{\left(\sum_{j\in C}\alpha_{jm}^{\theta_m/\theta_{rs}}e^{-\theta_m\beta x_j}\right)^{\frac{\theta_{rs}}{\theta_m}}}{\sum_{p=1}^{M}\left(\sum_{j\in C}\alpha_{jp}^{\theta_p/\theta_{rs}}e^{-\theta_p\beta x_j}\right)^{\frac{\theta_{rs}}{\theta_p}}} \tag{36}$$

where C_m is the choice set that includes all the routes that use link m, M is the number of links in Γ, the inclusive parameters are $\alpha_{im} = l_m/L_i$, and θ_m is the scale parameter of nest m, which is to be estimated.

- *A probit model*, which is taken as the "true" model that represents the true route choice behaviour. In this case, it is solved using a Monte Carlo simulation in which each link has a random utility $U_a = -l_a + \varsigma_a$, where $\varsigma_a \sim N(0, \lambda l_a)$. The probability of the route choice is calculated as the average from one million simulation cycles.

The scale parameters θ_{rs} for the logit-based models are assumed to be one. The route length is the only attribute in x_i and the corresponding parameter β is assumed to be 1 (notice we have a negative sign in front of β and therefore length is a disutility). The correction parameter $corr_i$ is assumed to be 1 for the C-logit model and -1 for the PSL and PSCL models. The paired scale parameter of the non-overlapping part is $\theta_{rg} = \sqrt{10/(10-l_a)}$ for the new model. All the nesting scale parameters in the CNL model are set at $\theta_m = 1.5$ or $\theta_m = 2$, identified as CNL-1.5 and CNL-2 in Fig. 4. The probit model uses the scale parameter λ whereas the MNL, PCL, CNL, PSL and C-logit models use the parameter θ_{rs}. To ensure consistency among different models, we assume that the variance of the logit perception errors is the same as that of the shortest route in the probit model, which is $\pi^2/6\theta_{rs}^2 = \lambda L_{\min}$, where $L_{\min}$ is the length of the shortest route in choice set C_{rs}.

(2) The formula and specification of the LK model used in "Application in a real road network" section: The utility of route i in the LK model is

$$U_i = V_i + \mathbf{F}_{(N \times N)} \mathbf{T}_{(N \times N)} \xi_{(N \times 1)} + \varepsilon_i, \tag{37}$$

where $\mathbf{F}_{(N \times N)}$ is the factor loadings matrix (N is the number of routes) and an element f_{ij} of $\mathbf{F}_{(N \times N)}$ is the length which path i overlaps with path j $(i, j \in C)$ and $\mathbf{T}_{(N \times N)} = diag(\lambda_1, \lambda_2, \ldots, \lambda_N)$ (λ_N is the covariance parameter, which is to be estimated). We assume that all parameters are the same $\lambda_{LK} = \lambda_1 = \lambda_2 = \cdots = \lambda_N$, $\xi_{(N \times 1)}$ is a vector of the IID $N(0, 1)$ variate, and ε_i is a vector of the IID Gumbel distributed variate.

References

Azevedo, J., Costa, M.S., Madeira, J.S., Martins, E.V.: An algorithm for the ranking of shortest paths. Eur. J. Oper. Res. **69**(1), 97–106 (1993). doi:10.1016/0377-2217(93)90095-5

Barton, R.R., Hearn, D.W.: Network aggregation in transportation planning models. Report DOT-TSC-RSPA-79-18. United States Department of Transportation (1979). https://trid.trb.org/view/89025

Bekhor, S., Ben-Akiva, M.E., Ramming, M.S.: Adaptation of logit kernel to route choice situation. Transp. Res. Rec. J. Transp. Res. Board **1850**, 78–85 (2007). doi:10.3141/1805-10

Bekhor, S., Prashker, J.N.: Stochastic user equilibrium formulation for generalized nested logit model. Transp. Res. Rec. **1752**, 84–90 (2001). doi:10.3141/1752-12

Bekhor, S., Toledo, T., Prashker, J.N.: Effects of choice set size and route choice models on path-based traffic assignment. Transportmetrica **4**(2), 117–133 (2008)

Ben-Akiva, M.E., Bierlaire, M.: Discrete choice methods and their applications to short term travel decisions. In: Hall, R.W. (ed.) Handbook of Transportation Science, pp. 5–33. Kluwer, Dordrecht, Netherlands (1999)

Bierlaire, M.: A theoretical analysis of the cross-nested logit model. Ann. Oper. Res. **144**(1), 287–300 (2006)

Bliemer, M.C.J., Bovy, P.H.L.: Impact of route choice set on route choice probabilities. Transp. Res. Rec. **2076**, 10–19 (2008). doi:10.3141/2076-02

Bovy, P.H.L., Bekhor, S., Prato, C.G.: The path size factor revisited: an alternative derivation. In: Proceedings of the 87th Transportation Research, Washington DC, USA (2008)

Boyles, S.D.: Bushed-based sensitivity analysis for approximating subnetwork diversion. Transp. Res. Part B Methodol. **46**(1), 139–155 (2012). doi:10.1016/j.trb.2011.09.004

Cascetta, E.: Transportation Systems Analysis: Models and Applications. Springer, New York (2009)

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A.: A modified logit route choice model overcoming route overlapping problems: specification and some calibration results for interurban networks. In: Transportation and traffic theory. Proceedings of the 13th international symposium on transportation and traffic theory, Lyon, France, 24–26 July 1996, pp 697–711. Elsevier, Oxford (1996)

Chen, A., Pravinvongvuth, S., Xu, X., Ryu, S., Chootinan, P.: Examining the scaling effect and overlapping problem in logit-based stochastic user equilibrium models. Transp. Res. Part A **46**, 1343–1358 (2012). doi:10.1016/j.tra.2012.04.003

Chu, C.: A paired combinatorial logit model for travel demand analysis. In: Proceedings of the Fifth World Conference on Transportation Research, pp 295–309. Ventura, CA, 1989

del Castillo, J.M.: A class of rum choice models that includes the model in which the utility has logistic distributed errors. Transp. Res. Part B Methodol. **91**, 1–20 (2016). doi:10.1016/j.trb.2016.04.022

Fosgerau, M., Frejinger, E., Karlstrom, A.: A link based network route choice model with unrestricted choice set. Transp. Res. Part B Methodol. **56**, 70–80 (2013). doi:10.1016/j.trb.2013.07.012

Frejinger, E., Bierlaire, M.: Capturing correlation with subnetworks in route choice models. Transp. Res. Part B Methodol. **41**, 363–378 (2007). doi:10.1016/j.trb.2006.06.003

Kazagli, E., Bierlaire, M., Flötteröd, G.: Revisiting the route choice problem: a modeling framework based on mental representations. In, vol. Technical report. Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne (2016)

Kitthamkesorn, S., Chen, A.: A path-size weibit stochastic user equilibrium model. Transp. Res. Part B Methodol. **57**, 378–397 (2013). doi:10.1016/j.trb.2013.06.001

Kitthamkesorn, S., Chen, A.: Unconstrained weibit stochastic user equilibrium model with extensions. Transp. Res. Part B Methodol. **59**, 1–21 (2014). doi:10.1016/j.trb.2013.10.010

Lai, X., Li, J., Li, Z.: A subpath-based logit model to capture the correlation of routes. Promet Traffic Transp. **28**(3), 235–245 (2016)

Li, B.: The multinomial logit model revisited: a semi-parametric approach in discrete choice analysis. Transp. Res. Part B Methodol. **45**(2), 461–473 (2011)

Li, J., Huang, Y., Lai, X.: Modeling stochastic route choice behaviors with equivalent impedance. Math. Probl. Eng. **2015**, 10 (2000)

McFadden, D., Train, K.: Mixed MNL models for discrete response. J. Appl. Econom. **15**(5), 447–470 (2000)

Prashker, J., Bekhor, S.: Investigation of stochastic network loading procedures. Transp. Res. Rec. J. Transp. Res. Board **1645**, 94–102 (1998). doi:10.3141/1645-12

Prato, C.G.: Route choice modeling: past, present and future research directions. J. Choice Model. **2**(1), 65–100 (2009). doi:10.1016/S1755-5345(13)70005-8

Prato, C.G., Bekhor, S.: Modeling route choice behavior: how relevant is the composition of choice set? Transp. Res. Rec. **2003**, 64–73 (2007). doi:10.3141/2003-09

Ramming, M.S.: Network knowledge and route choice. Massachusetts Institute of Technology (2002)

Sheffi, Y.: Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice Hall, London (1985)

Vovsha, P., Bekhor, S.: Link-nested logit model of route choice: overcoming route overlapping problem. Transp. Res. Rec. **1645**, 133–142 (1998). doi:10.3141/1645-17

Walker, J.: Extended discrete choice models: integrated framework, flexible error structures, and latent variables. MIT (2001)

Xie, C., Kockelman, K., Waller, S.T.: Maximum entropy method for subnetwork origin–destination trip matrix estimation. Transp. Res. Rec. J. Transp. Res. Board **2196**, 111–119 (2010). doi:10.3141/2196-12

Yao, J., Chen, A.: An analysis of logit and weibit route choices in stochastic assignment paradox. Transp. Res. Part B Methodol. **69**, 31–49 (2014). doi:10.1016/j.trb.2014.07.006

Jun Li is currently an Associate Professor in the School of Engineering, Sun Yat-sen University, specializing in network analysis and transportation economics. He received his PhD in Civil Engineering from Nagoya University in 2001.

Xinjun Lai is currently a research fellow specializing in discrete choice analysis and travel behaviour modelling. She received her PhD degree from Sun Yat-sen University in 2014.