Transportation Research Part E 127 (2019) 284–307

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

Bike network design problem with a path-size logit-based equilibrium constraint: Formulation, global optimization, and matheuristic

Haoxiang Liu[a], W.Y. Szeto[b,c,*], Jiancheng Long[a]

[a] School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei 230009, China
[b] Department of Civil Engineering, The University of Hong Kong, Hong Kong
[c] The University of Hong Kong Shenzhen Institute of Research and Innovation, Shenzhen, China

ARTICLE INFO

Keywords:
Bicycle network design
Global optimization method
Path-size logit
Matheuristic

ABSTRACT

This study focuses on the optimal network design problem of bike paths, which are on or adjacent to roadways but are physically separated from motorized traffic within the existing urban network. The problem seeks to maximize the total route utilities of cyclists and capture their actual route choice behavior using a path-size logit model. A mixed-integer nonlinear nonconvex model is developed for the problem and is reformulated and linearized into a mixed-integer linear program. The program is solved with a global optimization method and a matheuristic. Results are provided to illustrate the performance of these methods and the model properties.

1. Introduction

Cycling is a sustainable and environmentally friendly transportation mode that can decrease greenhouse gas emissions and reduce traffic congestion in urban areas. However, a lack of bicycle infrastructure is a great barrier to the promotion of cycling. The provision of bicycle facilities is effective in improving cycling safety and the cycling experience. System infrastructure design for highway traffic (e.g., private cars and trucks) and public transit (e.g., subways and buses) has received considerable attention in the literature from both researchers and policymakers, but fewer efforts have been focused on system infrastructure design for cycling, such as the design of bikeway networks. In general, budgets are limited, and the infrastructure is expensive. It is therefore important to carefully allocate funding to construct bicycle infrastructure within a road network to maximize its network performance.

Bicycle infrastructure and facilities include bike lanes that are separated from motorized traffic by white lines, cycle tracks (bicycle-exclusive paths physically separated from motor vehicles and pedestrians), bicycle boulevards, cycling bridges, and other special facilities for cycling. Cyclists' preferences for cycling facilities have been reported in recent studies. In Portland, Oregon, for example, about 50% of the kilometers cycled were associated with bicycle facilities, but they accounted for only 8% of the bikeable road network (Broach et al., 2012; Dill, 2009). Howard and Burns (2001) showed that cyclists might adjust their routes to use bicycle facilities. Reviews of the effects of various bicycle facilities on cyclist preferences and the relationship between bikeway networks and cycling levels were published by Pucher and Buehler (2008) and Buehler and Dill (2016).

This study specifically focuses on the development of modeling and solution methods for the network design of separated bike-exclusive paths (referred to hereafter as "bike paths"). Improvements in cyclists' safety have been associated with cycling along bike paths compared to cycling along comparable reference streets (Lusk et al., 2011), and the cycling experience and cycling levels have

* Corresponding author at: Department of Civil Engineering, The University of Hong Kong, Hong Kong.
E-mail address: ceszeto@hku.hk (W.Y. Szeto).

https://doi.org/10.1016/j.tre.2019.05.010
Received 19 July 2018; Received in revised form 8 April 2019; Accepted 12 May 2019
Available online 04 June 2019

improved after the construction of bike paths (Goodno et al., 2013; Lusk et al., 2011; Snizek et al., 2013). Moreover, Pucher and Buehler (2006, 2008) found that one key to a high level of cycling might be the supply of separate cycling facilities along heavy traffic roadways, which also implies that bike paths may help to enhance cycling levels. The bike path network design problem is important for improving the cycling level, cycling experience, and cyclists' safety.

The route choice behavior of cyclists is an important area related to the bike path network design problem because their behavior governs the bike flow pattern in a bicycle network and affects the final decision of where to build bike paths. Revealed preference survey data showed that cyclists' behavior was related to their preference for separate cycling facilities and not necessarily to the shortest distance and it could be well modeled by a path-size logit (PSL) model. Aultman-Hall et al. (1997) investigated 397 routes used by commuter cyclists in Guelph, Ontario, Canada, and reported that the routes selected by these cyclists were on average 0.4 km longer than the shortest available routes. Krizek et al. (2007) studied the travel behavior of cyclists in an urban trail corridor in Minneapolis, Minnesota, United States, and found that the cyclists' routes were on average 67% longer than the shortest available route to include the trail facility on their routes. Winters et al. (2010) compared the actual routes used by cyclists in metropolitan Vancouver, Canada, and the shortest path generated by a transportation model. They found that 75% of trips were within 10% of the shortest distance and that about 90% were within 25%, and, as expected, the cyclists detoured to use bike facilities. Some studies have presented detailed discrete-choice models to depict cyclists' route choice behavior. Hood et al. (2011) developed a PSL bicycle route-choice model based on the GPS data collected from smartphone users in San Francisco. Broach et al. (2012) developed a bicycle route-choice model based on the revealed preference GPS data from 164 cyclists in Portland, Oregon, United States. The model also takes the form of a PSL to account for overlapping route alternatives. These studies found that the utility of a route was correlated to multiple route attributes, especially bike facility types. Broach et al. (2012) reported that separate bike paths were the cyclists' priority preference.

Few studies have developed different models of the optimal designs of bicycle networks, although guidance regarding where to construct bicycle facilities is needed. These studies considered only the shortest route or deviation of the shortest route for bike demand in their design models, but they did not consider realistic cyclists' route choice behavior. Smith and Haghani (2012) developed a mixed-integer programming model to determine the locations of bicycle routes and bicycle lanes for an existing urban road network given an investment budget cap. They considered two objectives: minimizing the shortest route and maximizing the bicycle level of service between all origin-destination (OD) pairs. These objectives were combined using a weighted-sum approach. Duthie and Unnikrishnan (2014) presented a bike network design formulation to retrofit the current roadway infrastructure to achieve a minimum construction investment. They aimed to create bike routes between each OD pair, and the service levels of these routes were at least as good as a predetermined service level. The length of each route was constrained to a maximum acceptable upper bound, which was expressed as a function of the shortest route length. Some studies simply assumed that bike users followed Wardrop's first principle (Wardrop, 1952) to select routes rather than the more realistic discrete route-choice model calibrated by survey data (Broach et al., 2012; Hood et al., 2011). Because it is not always possible to construct high-quality off-road bike facilities, Mesbah et al. (2012) proposed an optimization framework for an on-road bike lane design in an existing urban road network. The benefit to bike demand due to the allocation of new bike lanes and the disbenefit to motorized traffic were both considered in their study. A bilevel optimization model was proposed to maximize the total travel distance on bike lanes and minimize the total travel time of cars in the upper-level program using a weighted-sum approach and to include the user equilibrium traffic assignment formulation for both bicycle and car demand in the lower-level program.

The car and transit network designs are two areas related to this study. In recent decades, far more research has been carried out in these areas than in bicycle network design, including continuous road network design (e.g., Abdulaal and LeBlanc, 1979; Szeto et al., 2015; Wang and Lo, 2010; Wang et al., 2014), discrete network design (e.g., Gao et al., 2005; LeBlanc, 1975; Liu and Wang, 2016; Long et al., 2010; Wang et al. 2013; Wang et al. 2015), mixed network design (e.g., Luathep et al., 2011), transit network design (e.g., An and Lo, 2015; Miandoabchi et al., 2012; Nayan and Wang, 2017; Pattnaik et al., 1998; Szeto and Jiang, 2014; Wu et al., 2015), and road pricing (e.g., Ekström et al., 2012; He et al., 2017; Lu et al., 2015; Szeto and Lo, 2006; Yang and Huang, 2005; Zhang et al., 2008). Recent reviews of these research areas were published by Yang and Bell (1998), Farahani et al. (2013), Kepaptsoglou and Karlaftis (2009), and Yang and Huang (2005). It is very important to capture the congestion effect (i.e., travel time to be an increasing function of the flow) in car and transit networks and their design. Moreover, vehicle flow is sensitive to travel time and travel cost (e.g., pricing) in the determination of routes. However, within bike networks, cyclists consider other factors in their route choices, such as safety and a preference for bike facilities. Furthermore, the congestion effect has not yet been found to be a significant factor for route choice in US cities (Broach et al., 2012; Hood et al., 2011). A new and realistic model framework is thus needed for the bicycle network design problem to incorporate cyclists' route choice behavior.

Global optimization algorithms have recently attracted attention from many researchers in the area of solution algorithms for network design problems (e.g., Ekström et al., 2013; Li et al., 2012; Liu and Wang, 2015; Luathep et al., 2011; Szeto et al., 2015; Wang and Lo, 2010; Wang et al. 2013; Wang et al. 2015), although most studies on global optimization algorithms focus on various car network design problems and very few focus on planning of bike facilities. The advantage of global optimization algorithms is that an exact solution to the problem can be obtained, which is significant from a practical point of view. However, their computational inefficiency is always an obstacle for the wide application of global optimization algorithms because, in practice, time is limited to obtain network design solutions. For example, transportation planning consultants have limited time to finish a network design project, and they must consider many different scenarios because the values for some of the parameters may not be certain. Liu and Wang (2015) presented a global optimization method for a continuous network design problem, in which a range-reduction technique was applied to reduce the feasible region and thus the computation time for the global optimization method. One advantage of this algorithm is that the exact solution must remain within the updated feasible region and be obtained definitely. However, the

computational efficiency of the global optimization algorithm remains far behind that of local optimization algorithms because the range-reduction technique itself is also a computational burdened. In this paper, a global optimization method based on mixed-integer linear programming is presented to solve the bicycle network design problem, and a novel matheuristic (i.e., a hybrid of a global optimization method and a heuristic) is then proposed to efficiently solve large bike network design problems. Inspired by the spirit of Liu and Wang (2015), the feasible region of the bike network design problem is tightened with a surrogate-model–based heuristic before application of the global optimization method introduced in this paper. Compared to the range-reduction technique, the surrogate-model–based method makes the bound update tractable. In this way, the novel method achieves a more accurate result and leads to a much lower computation time, according to the results presented in the section of numerical studies.

This study investigates a bike path network design problem. The problem is presented from the perspective of the transportation authority and is intended to maximize the total route utilities of all cyclists with consideration of their realistic route choice behavior. The total construction cost of a design is subject to a given government investment budget. This study differs from previous studies in four respects. First, the route utility that involves many attributes to cycling, such as route length, turn frequency, the slope of the route, and the presence of bicycle infrastructure, is adopted to improve the bicycle network design, rather than simply using the shortest route length between each OD pair, as in previous studies. Second, we propose a bilevel bike path network design model. A PSL model developed from real survey data to describe cyclists' route choice behavior, specifically for cities in the United States (Broach et al., 2012), is incorporated as the lower-level model, which has not been used in the bicycle network design literature. The PSL model considers the factors that affect cyclists' choice of routes and the overlap between routes. Such inclusion leads to a more realistic bike path network design model. Moreover, although the introduction of bike paths to the existing urban network improves route utility and affects cyclists' route choice, their choices also affect the construction decision. This key interaction is captured in the proposed bilevel bike path network design model. Third, due to the difficulty of solving a bilevel model with multivariate nonlinear functions, no global optimization methods have yet been presented in the literature of bicycle network design, although such methods have recently been used in road network design. The proposed model is reformulated into an equivalent model, to which an efficient two-dimensional linearization method can be directly applied. Fourth, computational efficiency is always the biggest issue for wide application of global optimization methods. In this paper, a novel matheuristic is proposed to greatly improve the computational efficiency of large network applications and provide good solutions. The matheuristic is developed by incorporating a surrogate-model–based method into the proposed global optimization method to tighten the feasible region of variables and accelerate the computation process.

The remainder of this paper is organized as follows: Section 2 provides the model formulation, Section 3 depicts the exact solution method and the novel matheuristic, Section 4 presents the numerical studies, and Section 5 concludes the paper.

2. Model formulation

We consider a network with multiple OD pairs, where W represents the set of cycling OD pairs. All roadways (i.e., links) in the network are currently suitable for or allow cycling; roadways that are not suitable for or allow cycling (e.g., freeways or roads on very steep slopes) have already been removed. It is also assumed that not all roadways that are suitable for cycling are also suitable for construction of an exclusive bike path. For example, the buildings along both sides of some urban roads are too close to the roadside to allow space for construction of a bike path. Thus, A denotes the set of all links in the network that are suitable for cycling and $\widetilde{A}$ denotes the set of candidate links along which bike paths can be constructed. Clearly, $\widetilde{A} \subset A$. The bike paths are not required to be connected. Moreover, the bike paths may not be identical to cyclists' routes. (For clarity and consistency, we use the term "route" to describe the course taken by a cyclist to get from his or her starting point to his or her destination, and we use the term "path" either to refer to a segregated bike-exclusive track or in the term "path-size factor" for the PSL model.) The cycling demand d^w of each OD pair w is assumed to be fixed, and it is assumed that cyclists choose their routes with the perceived maximal travel utility. The original route utility $U^0_{p,w}$ includes the characteristics of the current road network, such as roadway slopes and existing bicycle facilities. We also assume that the transportation authority plans the bicycle network within a given budget, which is common in practice.

The following notations are used.

Variables

f_p^w	The flow on route p between OD pair w.
$\mathbf{f}$	The vector of route flow, $\mathbf{f} = [f_p^w]$.
U_p^w	The cyclist's utility of choosing route p between OD pair w.
$\mathbf{U}$	The vector of route utilities, $\mathbf{U} = [U_p^w]$.
y_a	The binary variable for bike path construction on link a, where $y_a = 1$ if a bike path is introduced on link a and $y_a = 0$ otherwise.
$\mathbf{y}$	The vector of binary variables, $\mathbf{y} = [y_a]$.
Pr_p^w	The probability of using route p between OD pair w.
$\mathbf{Pr}$	The vector of route selection probabilities, $\mathbf{Pr} = [\mathrm{Pr}_p^w]$.
x_a	The bicycle flow on link a, $\mathbf{x} = [x_a]$.
PS_p^w	The path-size factor of route p between OD pair w that is used in the PSL model, which provides the information on the degree to which route p overlaps with other routes between the same OD pair w. It is used to reduce the utility of route$p$ that overlaps with other routes.

Parameters and sets

A	The set of links of the network.
$\widetilde{A}$	The set of candidate links for the introduction of bike paths.

A^p	The set of links on route p.
W	The set of OD pairs.
P^w	The set of available routes between OD pair w.
l_a	The length of link a.
B	The budget for bicycle network construction.
τ	The cost for bike path construction per unit length.
L_p	The length of route p.
$\delta_{a,p}^w$	An indicator that relates links to routes, where $\delta_{a,p}^w = 1$ if link a is on route p between OD pair w and $\delta_{a,p}^w = 0$ otherwise.
$U_{p,w}^0$	The original utility of route p between OD pair w before planning.
φ	The coefficient associated with the factor—the proportion of the total length of bike paths on a feasible route to the length of that route—in the route utility functions.
θ	The scaling parameter of the path-size factor.
d^w	Cycling demand between OD pair w.

The studied bike path network design problem is formulated from the perspective of the transportation authority to maximize social welfare (the total route utilities of all cyclists) within a given investment budget. The decision is to determine which links should have bike paths. The cyclists' route choice behavior, which is presented in a form of the PSL model according to Broach et al. (2012), is considered in the problem formulation to make the bike path network planning decision more realistic. Thus, we propose the following mathematical programming model for the problem.

(P0)

$$\min_{\mathbf{y,f,U,Pr}} Z = -\sum_{w\in W}\sum_{p\in P^w} f_p^w U_p^w \tag{1}$$

subject to

$$\sum_{a\in\widetilde{A}} y_a l_a \tau \leqslant B \tag{2}$$

$$y_a \in \{0, 1\} \quad \forall\, a \in \widetilde{A} \tag{3}$$

$$f_p^w = d^w \mathrm{Pr}_p^w \quad \forall\, p \in P^w,\, w \in W \tag{4}$$

$$\mathrm{Pr}_p^w = \frac{\exp(U_p^w + \theta \ln(PS_p^w))}{\sum_{q\in P^w}\exp(U_q^w + \theta \ln(PS_q^w))} \quad \forall\, p \in P^w,\, w \in W, \text{and} \tag{5}$$

$$U_p^w = U_{p,w}^0 + \varphi \sum_{a\in A} \frac{l_a}{L_p} y_a \delta_{a,p}^w \quad \forall\, p \in P^w,\, w \in W \tag{6}$$

where

$$PS_p^w = \sum_{a\in A^p} \frac{l_a}{L_p} \frac{1}{\sum_{q\in P^w}\delta_{a,q}^w} \quad \forall\, p \in P^w,\, w \in W \tag{7}$$

Constraint (2) ensures that the construction cost is not greater than the investment budget available. Eq. (3) is the binary constraint for the decision variable y_a for bike path construction on candidate link a, which takes the value of 1 if a bike path is constructed on link a and 0 otherwise. Eq. (4) means that the demand is allocated to each route based on its selection probability.

Eqs. (5)–(7) describe the cyclists' route choice behavior. Eq. (5) states that the probability of selecting route p between OD pair w is given by the PSL model (Ben-Akiva and Bierlaire, 1999; Frejinger and Bierlaire, 2007). The PSL model states that the probability of a cyclist choosing a route between an OD pair relates to the utility of each route between this OD pair (i.e., U_p^w) and a correction term for each route to account for the similarities among the routes between the same OD pair, that is, $\theta \ln(PS_p^w)$. The route utility U_p^w captures the factors that affect cyclists' selection of routes, such as route length, turn frequency, the slope of the route considered, and the presence of bicycle infrastructure (Broach et al., 2012). A shorter route length, a lower turn frequency, a lesser slope, and greater provision of bicycle infrastructure lead to greater route utility and greater probability that the route will be used.

In this study, we consider only the effects of the change in a bicycle network on the change in the utility function value. Moreover, we assume that other factors that affect cyclists' route choice remain constant and are already included in the original route utility. It is also assumed that cyclists ride on roads before the construction of exclusive bike paths. Thus, the utility function of route p between OD pair w can be expressed as Eq. (6), where the first term $U_{p,w}^0$ represents the original utility of route p between OD pair w before bike paths are constructed on the route and the second term represents the additional utility after bike paths are completed. The original route utility $U_{p,w}^0$ is a constant (i.e., independent of bike flow) and includes all factors that affect route choice, such as route length, turn frequency, and slope, except the presence of a bike path. Once the information of the initial network is given, $U_{p,w}^0$ can be calculated with the utility function of Hood et al. (2011) and Broach et al. (2012). The change of route utility reflects only the benefits of the construction of bike paths and is mathematically expressed as $\varphi \sum_{a\in A} \frac{l_a}{L_p} y_a \delta_{a,p}^w$ (i.e., the second term of U_p^w in Eq. (6), which represents the contribution of the increase in the proportion of the route that includes a bike path to the route utility).

The path-size factor adopted in the PSL model in Eq. (5) is expressed as Eq. (7). The term inside the summation sign in Eq. (7) is the relative length of each link (with respect to the route length) shared by other routes in the same OD pair divided by the number of

routes in the same OD pair that pass through this link. The path-size factor is fixed once the set of alternative routes between the OD pair, the network geometry, and the length of each link are given. The path-size factor provides the information on the degree to which the route overlaps with other routes between the same OD pair. In particular, $PS_p^w = 1$ if route p has no overlapping link with other routes, and $PS_p^w = 0.5$ if two routes completely overlap.

The full model formulation is given by (1)–(6), which is a mixed-integer nonlinear nonconvex program that is difficult to solve. In Section 3.1, we propose a global optimization method and a matheuristic to solve this model.

For the purpose of analysis in Section 4, we also present an expression of the bike flow on link $a \in A$ as follows:

$$x_a = \sum_{w \in W} \sum_{p \in P^w} d^w \Pr_p^w \delta_{a,p}^w \quad \forall\, a \in A \tag{8}$$

Note that the route set in the PSL model is predetermined in our network design formulation. For route-based network design models, it is common to generate a route set before solving the problem. In fact, route set generation is a traditional problem, and many researchers have studied the route set generation method for route-based network design problems (e.g., Bekhor et al., 2001, 2006; Bekhor and Toledo, 2005; Broach et al., 2010; Dial, 1971; Han, 2007). Hence, the implicit assumption of a given route set does not affect the application of route-based network design models (including ours) to large networks.

3. Solution methods

This section presents two solution methods, including a global optimization method and an efficient novel matheuristic. The first method theoretically guarantees that an exact solution will be obtained for the linearized problem, whereas the latter greatly improves the computational efficiency for the solution of large network instances by embedding a surrogate-model–based heuristic within the first method.

3.1. Global optimization method

Due to the nonlinear nonconvex characteristic of the presented bicycle network design model, a global optimum is difficult to obtain. An efficient mixed-integer linearization technique is used to solve the model for (P3) to obtain a nearly global optimum, which is shown to be quite near the exact solution in the numerical study section. The main idea of this method is to transform the original model with multivariate nonlinear functions into an equivalent form with two-variable nonlinear functions that can be directly and efficiently linearized with a two-dimensional linearization technique. The equivalent model is then linearized in a piecewise manner by first linearizing the nonconvex functions in a piecewise manner and applying the logarithmic branching convex combination modeling technique to formulate the piecewise linear constraints (Vielma et al., 2010). Finally, the reformulated model is solved with a classic algorithm for mixed-integer linear programs (MILPs) or with a commercial solver to its global optimum.

3.1.1. Model transformation

In this subsection, the original model is reformulated into an equivalent nonlinear nonconvex program that can be linearized directly and efficiently with a linearization technique. Two nonlinear terms in the original model require linearization: the probability function on the right-hand side of Eq. (5) and the total utilities in the objective function.

Direct linearization of the probability function is quite complicated and inefficient because the nonlinear probability function of each specific route involves multiple independent variables (i.e., the utilities of each alternative route). Hence, it is more convenient and efficient if Eq. (5) is transformed into functions with two or fewer independent variables. To this end, a new continuous variable $\alpha^w \in \mathbb{R}$ is introduced for each OD pair $w \in W$ to replace $\frac{1}{\sum_{q \in P^w} \exp(U_q^w + \theta \ln(PS_q^w))}$, and according to Proposition 1, Eq. (5) can be reformulated as follows:

$$\Pr_p^w = \alpha^w (PS_p^w)^\theta \exp(U_p^w) \quad \forall\, p \in P^w,\, w \in W \text{ and} \tag{9}$$

$$\sum_{p \in P^w} \Pr_p^w = 1 \quad \forall\, w \in W \tag{10}$$

Proposition 1. *For $\alpha^w \in \mathbb{R}$, $w \in W$, constraints (9) and (10) are equivalent to Eq. (5).*

Proof.

(a) Eq. (5) ⇒ constraints (9) and (10): $\forall\, w \in W$, $\sum_{p \in P^w} \Pr_p^w = \sum_{p \in P^w} \frac{\exp(U_p^w + \theta \ln(PS_p^w))}{\sum_{q \in P^w} \exp(U_q^w + \theta \ln(PS_q^w))} = \frac{\sum_{p \in P^w} \exp(U_p^w + \theta \ln(PS_p^w))}{\sum_{q \in P^w} \exp(U_q^w + \theta \ln(PS_q^w))} = 1$, and hence constraint (10) is satisfied. Moreover, for each $p \in P^w$, we rewrite Eq. (5) to $\Pr_p^w = \frac{1}{\sum_{q \in P^w} \exp(U_q^w + \theta \ln(PS_q^w))} \exp(U_p^w + \theta \ln(PS_p^w))$, and let $\alpha^w = \frac{1}{\sum_{q \in P^w} \exp(U_q^w + \theta \ln(PS_q^w))}$. We then have constraint (9) and $\alpha^w \in \mathbb{R}$, $w \in W$.

(b) constraints (9) and (10) ⇒Eq. (5): $\forall\, w \in W$, substituting constraint (9) into constraint (10), we have $\sum_{p \in P^w} \Pr_p^w = \alpha^w \sum_{p \in P^w} (PS_p^w)^\theta \exp(U_p^w) = 1$. From this expression, we solve for α^w and get $\alpha^w = \frac{1}{\sum_{p \in P^w} (PS_p^w)^\theta \exp(U_p^w)} = \frac{1}{\sum_{p \in P^w} \exp(U_p^w + \theta \ln(PS_p^w))}$. Substitution of this expression into constraint (9) leads to Eq. (5). This ends the

proof. □

Note that in constraint (9), the route selection probability is only a function of its own utility U_p^w and α^w, which makes it easy to linearize constraints (9) and (10).

As for the objective function, by substituting Eqs. (4) and (9) into the objective (1), we obtain the new objective:

$$\min Z_1 = -\sum_{w\in W} d^w \sum_{p\in P^w} \alpha^w (PS_p^w)^\theta \exp(U_p^w) U_p^w \tag{11}$$

The advantage of this transformation is that the total utilities now depend also upon the route utility U_p^w and α^w, which are the same as the route selection probability function. In this manner, the number of additional variables and constraints in the linearization process can be greatly reduced, which is described in the next subsection. Moreover, the route flow f_p^w and constraint (4) can both be removed from the model because they are already implicit in the objective function. The size of the model formulation is thus further reduced.

Define $\boldsymbol{\alpha} = [\alpha^w]$. The equivalent problem can be formulated as follows:

(P1)

$$\min_{\mathbf{y,U,Pr},\boldsymbol{\alpha}} Z_1 = -\sum_{w\in W} d^w \sum_{p\in P^w} \alpha^w (PS_p^w)^\theta \exp(U_p^w) U_p^w \tag{12}$$

subject to

$$\sum_{a\in\widetilde{A}} y_a l_a \tau \leqslant B \tag{13}$$

$$y_a \in \{0, 1\} \quad \forall\, a \in \widetilde{A} \tag{14}$$

$$\sum_{p\in P^w} \mathrm{Pr}_p^w = 1 \quad \forall\, w \in W \tag{15}$$

$$\mathrm{Pr}_p^w = \alpha^w (PS_p^w)^\theta \exp(U_p^w) \quad \forall\, p \in P^w,\, w \in W, \text{and} \tag{16}$$

$$U_p^w = U_{p,w}^0 + \varphi \sum_{a\in A} \frac{l_a}{L_p} y_a \delta_{a,p}^w \quad \forall\, p \in P^w,\, w \in W \tag{17}$$

where

$$PS_p^w = \sum_{a\in A^p} \frac{l_a}{L_p} \frac{1}{\sum_{q\in P^w} \delta_{a,q}^w} \quad \forall\, p \in P^w,\, w \in W \tag{18}$$

For ease of linearization, we define two new variables, $R_O^{w,p}$ and $R_{\mathrm{Pr}}^{w,p}$, such that $R_O^{w,p} = \alpha^w (PS_p^w)^\theta \exp(U_p^w) U_p^w$, $R_{\mathrm{Pr}}^{w,p} = \alpha^w (PS_p^w)^\theta \exp(U_p^w)$, and we remove the variable Pr_p^w. Moreover, we define $\mathbf{R_O} = [R_O^{w,p}]$ and $\mathbf{R_{Pr}} = [R_{\mathrm{Pr}}^{w,p}]$. We then have another reformulated problem (P2) that is equivalent to the original problem (P0), with the last two constraints to be linearized. (P2) is expressed as follows:

(P2)

$$\min_{\mathbf{y,U},\boldsymbol{\alpha},\mathbf{R_O},\mathbf{R_{Pr}}} Z_2 = -\sum_{w\in W} d^w \sum_{p\in P^w} R_O^{w,p} \tag{19}$$

subject to

$$\sum_{a\in\widetilde{A}} y_a l_a \tau \leqslant B \tag{20}$$

$$y_a \in \{0, 1\} \quad \forall\, a \in \widetilde{A} \tag{21}$$

$$\sum_{p\in P^w} R_{\mathrm{Pr}}^{w,p} = 1 \quad \forall\, w \in W \tag{22}$$

$$U_p^w = U_{p,w}^0 + \varphi \sum_{a\in A} \frac{l_a}{L_p} y_a \delta_{a,p}^w \quad \forall\, p \in P^w,\, w \in W \tag{23}$$

$$R_{\mathrm{Pr}}^{w,p} = \alpha^w (PS_p^w)^\theta \exp(U_p^w) \quad \forall\, p \in P^w,\, w \in W, \text{and} \tag{24}$$

$$R_O^{w,p} = \alpha^w (PS_p^w)^\theta \exp(U_p^w) U_p^w \quad \forall\, p \in P^w,\, w \in W \tag{25}$$

where

$$PS_p^w = \sum_{a\in A^p} \frac{l_a}{L_p} \frac{1}{\sum_{q\in P^w} \delta_{a,q}^w} \quad \forall\, p \in P^w,\, w \in W \tag{26}$$

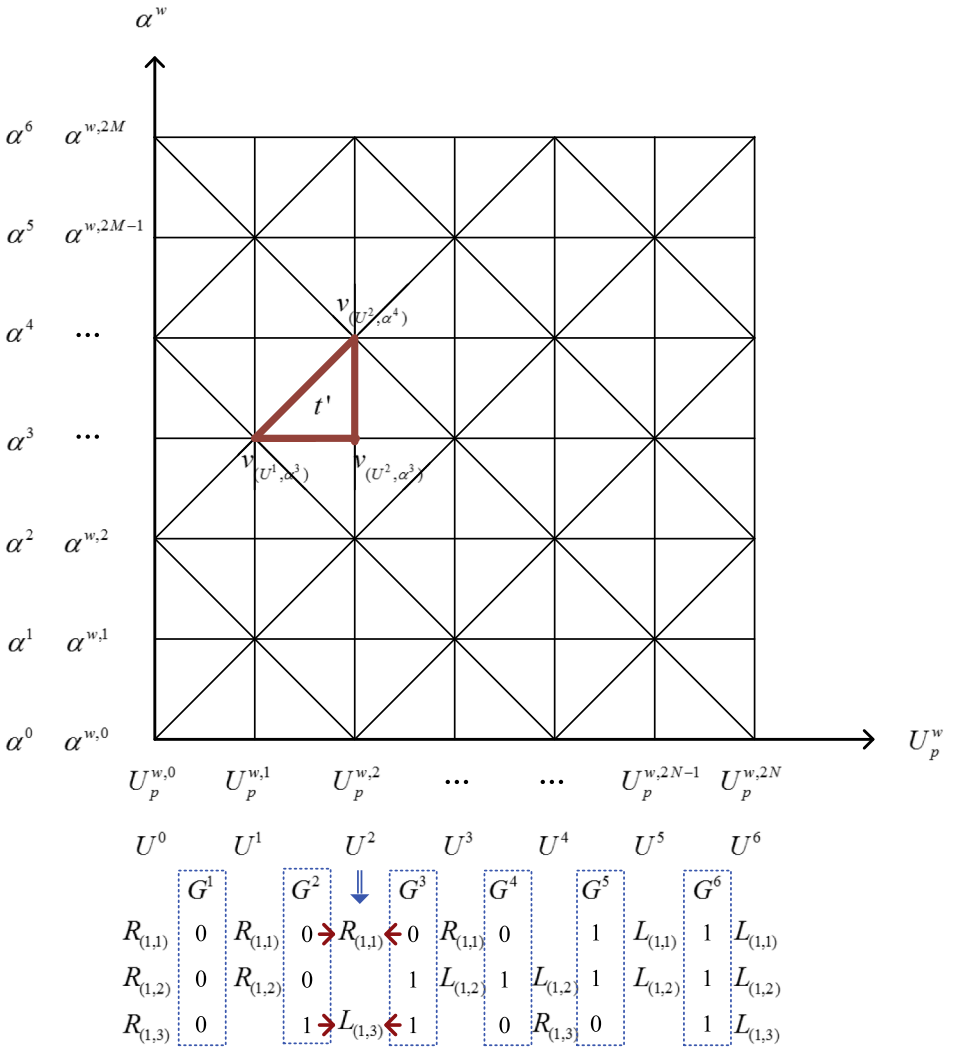


Fig. 1. Example of the compatible J_1 triangulation.

3.1.2. Nonlinear function linearization

Note that the last two constraints, Eqs. (24) and (25), are the only nonlinear constraints in problem (P2) due to the nonlinear functions on the right-hand sides and can be linearized and reformulated in a piecewise manner into mixed-integer linear constraints. Basically, to linearize a two-dimensional nonlinear function in a piecewise manner, the bounded feasible region is divided into a set of polytopes (i.e., triangles in the two-dimensional case), which follows the J_1 or *Union Jack* triangulation (Todd, 1977), and a flat plane formed by the nearest three corner points of a triangle is then used to approximate the curved surface of the original nonconvex function in this small triangle. Many studies on modeling nonconvex piecewise-linear functions as mixed-integer linear constraints have been published. In this study, the logarithmic branching convex combination model (Vielma and Nemhauser, 2011; Vielma et al., 2010), denoted by *Log*, is used to obtain the linearized right-hand sides of Eqs. (24) and (25). The characteristics of the Log model are as follows. First, it has fewer continuous variables than other linearization models. Second, it greatly reduces the number of discrete variables and constraints required by using the property of the Gray code and a logarithmic number of dichotomies to detect the active polytope $t \in T$. Here, the *active polytope* refers to the polytope that contains the global optimum. These characteristics ensure that the size of the resultant linearization model remains tractable even with a fine partition scheme.

Before the linearization model is proposed, some new notations and definitions must be introduced, and Eq. (24) is used as an example of the linearization process. The feasible regions of U_p^w and α^w are denoted as $[U_p^{w,\min}, U_p^{w,\max}]$ and $[\alpha^{w,\min}, \alpha^{w,\max}]$, respectively. The region $[U_p^{w,\min}, U_p^{w,\max}]\times[\alpha^{w,\min}, \alpha^{w,\max}]$, which is the bounded domain of the function $\alpha^w(PS_p^w)^\theta\exp(U_p^w)$, is divided into a set of polytopes T. To be compatible with the J_1 triangulation, the feasible region of U_p^w is divided into $2N$, $N \in \mathbb{Z}^+$ intervals, whereas the feasible region of α^w is divided into $2M$, $M \in \mathbb{Z}^+$ intervals. Fig. 1 shows the example triangulation adopted in this paper and an example polytope (bounded by thick lines). The horizontal axis is the U_p^w axis, and the vertical axis is the α^w axis. Let $V(T)$ be the set

of corner points of all polytopes T in the domain, and let $v \in V(T)$ be a corner point that may be associated with more than one triangle. (v_x, v_y) denotes the coordinates of a corner point v on the $U_p^w\alpha^w$-plane, and $R_{\mathrm{Pr}}^{w,p}(v_x, v_y)$ is the real value of $R_{\mathrm{Pr}}^{w,p}$ at (v_x, v_y). Note that (v_x, v_y) and $R_{\mathrm{Pr}}^{w,p}(v_x, v_y)$ are both known once T is fixed. λ_v is an additional continuous variable that represents the weight of each corner point v in an active polytope. To locate the active polytope t by adopting only $\kappa = \lceil Log_2\ |T| \rceil$ vectors of binary variables $\{\gamma_k\}_{k\in K}$, a series of dichotomies $\{L_{(k)}, R_{(k)}\}_{k\in K}$ is used, where K is an index set defined in the following text, and $L_{(k)}$ and $R_{(k)}$ are disjoint sets of corner points for each $k \in K$. By using the dichotomies, the corner points of every polytope t' can be located by $V(t') = \cap_{k\in K}(V(T)\backslash H_k)$, where $H_k = L_{(k)}$ or $H_k = R_{(k)}$.

The index set K contains two subsets ($K = K_1 \cup K_2$). The first is represented by $K_1 = \{(kx, ky)|kx \in \{1, ...,D\}, ky \in \{1, ...,\lceil \log_2 N_{kx}\rceil\}\}$, where D is the number of dimensions in the domain to be linearized, and for Eq. (24), $D = 2$; kx is used to define the axis. $kx = 1$ refers to the first axis (i.e., the U_p^w-axis). $kx = 2$ refers to the second axis (i.e., the α^w-axis). N_{kx} is the number of intervals on the kxth axis (i.e., $N_{kx} = 2N$, $N \in \mathbb{Z}^+$ when $kx = 1$, and $N_{kx} = 2M$, $M \in \mathbb{Z}^+$ when $kx = 2$). ky is used to define the bit position considered in the Gray code (Breckman, 1956) for a specific interval of the axis. For each $k = (kx, ky) \in K_1$, $L_{(kx,ky)} = \{v \in V|v_{kx} \in O(ky, 1)\}$ and $R_{(kx,ky)} = \{v \in V|v_{kx} \in O(ky, 0)\}$, where v_{kx} is the coordinate of the kxth axis, and $O(l, b) = \{n \in \{1, ...,N_{kx}\}|(n = 1 \text{ and } G_l^n = b) \text{ or } (G_l^n = b \text{ and } G_l^{n+1} = b) \text{ or } (n = N_{kx} \text{ and } G_l^n = b)\}$.$G^n \in \{0, 1\}^{\lceil \log_2 N_{kx}\rceil}$, $n \in \{1, 2, ...,N_{kx}\}$ refers to the Gray code for an interval n along the kxth axis. G_l^n is the lth bit of G^n. With the property of the Gray code, which is only one bit different from two adjacent codes, the number of binary variables can be reduced. Here, each interval index $n \in \{1, 2, ...,N_{kx}\}$ is represented by G^n.

The second subset K_2 is denoted by $K_2 = \{(kx, ky)|kx, ky \in \{1, ...,D\} \text{ and } kx < ky\}$. For each $k = (kx, ky) \in K_2$, $L_{(kx,ky)} = \{v \in V|v_{kx} \text{ is even and } v_{ky} \text{ is odd}\}$ and $R_{(kx,ky)} = \{v \in V|v_{kx} \text{ is odd and } v_{ky} \text{ is even}\}$, which are applied to distinguish the two polytopes in the same rectangle in this two-dimensional case. Fig. 1 gives an example of the use of the Gray code and dichotomies to locate an active polytope, which is marked in the figure and given below.

Consider a given path p between a specific OD pair w. For the sake of simplicity, the superscript w and the subscript p of U_p^w and α^w are omitted, and instead, the superscript of U and α refer to the index of breakpoints. In this example, seven breakpoints including the two endpoints are adopted on each of the two axes (i.e., $N = M = 3$). Hence, for the first subset K_1 and for each axis, a series of Gray codes with a length of $\lceil \log_2(6)\rceil = 3$ is needed to form the dichotomies. Taking the U-axis as an example (i.e., for $kx = 1$), each Gray code is unique and prespecified for an interval. By using the same bit—for example, the second bit (i.e., when the index $ky = 2$)—of two adjacent Gray codes, all corner points with the same U coordinate are classified as $R_{(1,2)}$ if both second bits of the two Gray codes equal 0, are classified as $L_{(1,2)}$ if they equal 1, and are not assigned to any set otherwise. Each axis has $\lceil \log_2 N_{kx}\rceil$ number of dichotomies for the subset K_1, which equals the length of the Gray codes for this axis. In this example, there are three dichotomies for the U-axis; that is, $\{L_{(1,1)}, R_{(1,1)}\}$, $\{L_{(1,2)}, R_{(1,2)}\}$, and $\{L_{(1,3)}, R_{(1,3)}\}$. It should be noted that the corner points with $U = U^0$ or U^6 only relate to one Gray code, whereas the corner points with $U = U^1$, U^2, U^3, U^4, or U^5relate to two adjacent Gray codes. Taking U^2 as an example, two adjacent Gray codes G^2 and G^3 with both the first bit equal to zero and the third bit equal to one lead to corner points $v_{(U^2,\alpha)} \in R_{(1,1)}$ and $v_{(U^2,\alpha)} \in L_{(1,3)}$. An active interval on the U-axis can be uniquely located by an intersection of $\{L_{(1,ky)}, R_{(1,ky)}\}_{(1,ky)\in K_1}$. Taking the interval between U^1 and U^2 as an example, the interval can be activated by locating two adjacent breakpoints $(V(T)\backslash L_{(1,1)})_{(1,1)\in K_1} \cap (V(T)\backslash L_{(1,2)})_{(1,2)\in K_1} \cap (V(T)\backslash R_{(1,3)})_{(1,3)\in K_1}$. For the α-axis, that is, when $kx = 2$, another series of dichotomies $\{L_{(2,ky)}, R_{(2,ky)}\}_{(2,ky)\in K_1}$ can be similarly formulated. The interval $[\alpha^3, \alpha^4]$ can be activated by $(V(T)\backslash L_{(2,1)})_{(2,1)\in K_1} \cap (V(T)\backslash R_{(2,2)})_{(2,2)\in K_1} \cap (V(T)\backslash L_{(2,3)})_{(2,3)\in K_1}$. Thus far, the rectangle between $[U^1, U^2]$ and $[\alpha^3, \alpha^4]$ is active. To further locate the marked polytope, the dichotomies related to the second subset K_2 are needed. The two-dimensional case has only one dichotomy $\{L_{(1,2)}, R_{(1,2)}\}_{(1,2)\in K_2}$, where all corner points with $U = U^0$, U^2, U^4 or U^6 and $\alpha = \alpha^1$, α^3 or α^5 belong to the set $L_{(1,2)}$ and all corner points with $U = U^1$, U^3 or U^5 and $\alpha = \alpha^0$, α^2, α^4 or α^6 belong to the set $R_{(1,2)}$. By using $(V(T)\backslash R_{(1,2)})_{(1,2)\in K_2}$, the corner points in the active rectangle but not in the marked polytope can be eliminated and the polytope t' can be activated by

$$V(t') = (V(T)\backslash L_{(1,1)})_{(1,1)\in K_1} \cap (V(T)\backslash L_{(1,2)})_{(1,2)\in K_1} \cap (V(T)\backslash R_{(1,3)})_{(1,3)\in K_1}$$
$$\cap (V(T)\backslash L_{(2,1)})_{(2,1)\in K_1} \cap (V(T)\backslash R_{(2,2)})_{(2,2)\in K_1} \cap (V(T)\backslash L_{(2,3)})_{(2,3)\in K_1} \cap (V(T)\backslash R_{(1,2)})_{(1,2)\in K_2}.$$

The value $R_{\mathrm{Pr}}^{w,p}$ of a point located in the polytope can then be approximated by the weighted sum of the function values at the three corner points in the active polytope.

Eq. (24) can be approximated with the following set of mixed-integer constraints:

$$\sum_{v\in V(T)} \lambda_v \cdot (v_x, v_y) = (U_p^w, \alpha^w) \tag{27}$$

$$\sum_{v\in V(T)} \lambda_v \cdot R_{\mathrm{Pr}}^{w,p}(v_x, v_y) = \widehat{R}_{\mathrm{Pr}}^{w,p} \tag{28}$$

$$\lambda_v \geqslant 0 \quad \forall\, v \in V(T) \tag{29}$$

$$\sum_{v\in V(T)} \lambda_v = 1 \tag{30}$$

$$\sum_{v\in L(k)} \lambda_v \leqslant \gamma_k \tag{31}$$

$$\sum_{v \in R(k)} \lambda_v \leqslant 1 - \gamma_k, \text{and} \tag{32}$$

$$\gamma_k \in \{0, 1\} \quad \forall\, k \in K \tag{33}$$

where $\widehat{R}_{\mathrm{Pr}}^{w,p}$ is the approximated value of $R_{\mathrm{Pr}}^{w,p}$ in (P2).

Eq. (27) indicates that the point (U_p^w, α^w) can be expressed as the weighted sum of the coordinates of the corner points of the active polytope. Eq. (28) illustrates that the value of the variable $R_{\mathrm{Pr}}^{w,p}$ can be approximated with the weighted sum of the corresponding $R_{\mathrm{Pr}}^{w,p}(v_x, v_y)$ value at the corner points of the active polytope with the same set of weights as in Eq. (27). The weight related to each corner point is non-negative, and the sum of the weights of all corner points is 1, which are expressed by conditions (29) and (30), respectively. By using the series of dichotomies, as shown in the small example above, constraints (31) and (32) ensure that only one active polytope exists in the feasible region—that is, all positive weights are associated with the same active polytope; the number of these constraints equals $\sum_{kx \in \{1,\ldots,D\}} 2[D \cdot \lceil \log_2 N_{kx} \rceil + D(D-1)/2]$ and depends on $\lceil \log_2 N_{kx} \rceil$ only when D is fixed. Condition (33) is the binary constraint for γ_k, $k \in K = K_1 \cup K_2$, which means that one binary variable corresponds to a dichotomy $\{L_k, R_k\}$. Conditions (27)–(33) are an approximation for constraint (24). Proposition 2 proves that the approximation approaches the original continuous constraint when $2N \to \infty$ and $2M \to \infty$.

Proposition 2. *For the given feasible regions* $[U_p^{w,\min}, U_p^{w,\max}]$ *and* $[\alpha^{w,\min}, \alpha^{w,\max}]$, *when the numbers of intervals* $2N \to \infty$ *and* $2M \to \infty$, *the mixed-integer constraints* (27)–(33) *approach constraint* (24).

Proof. Without loss of generality in this two-dimensional case, for a specific route p that belongs to OD pair w, we consider an active lower polytope in the divided domain with its three vertices (U_1, α_1), (U_2, α_1), and (U_2, α_2), the corresponding values of $R_{\mathrm{Pr}}^1 = R_{\mathrm{Pr}}(U_1, \alpha_1)$, $R_{\mathrm{Pr}}^2 = R_{\mathrm{Pr}}(U_2, \alpha_1)$, and $R_{\mathrm{Pr}}^3 = R_{\mathrm{Pr}}(U_2, \alpha_2)$, and the associated continuous variables λ_1, λ_2, and λ_3, in which the superscript and the subscript indicate that a specific route and a specific OD pair respectively are omitted for the sake of simplicity. Thus, we have $U_2 - U_1 = \Delta U = \frac{U_p^{w,\max} - U_p^{w,\min}}{2N}$ and $\alpha_2 - \alpha_1 = \Delta\alpha = \frac{\alpha^{w,\max} - \alpha^{w,\min}}{2M}$. For a general point (U^*, α^*) in this active polytope, the mixed-integer constraints have the following relationship:

$$\lambda_1 U_1 + \lambda_2 U_2 + \lambda_3 U_2 = U^* \text{ and (a)}$$
$$\lambda_1 \alpha_1 + \lambda_2 \alpha_1 + \lambda_3 \alpha_2 = \alpha^*. \text{ (b)}$$

Because $\lambda_1 + \lambda_2 + \lambda_3 = 1$ (*c*),
by solving the three linear groups of equations (a)–(c), we can obtain $\lambda_2 = \frac{U^* - U_1}{\Delta U} - \frac{\alpha^* - \alpha_1}{\Delta\alpha}$ and $\lambda_3 = \frac{\alpha^* - \alpha_1}{\Delta\alpha}$.
Let R_{Pr}^* be the real value of R_{Pr} at (U^*, α^*). The distance between R_{Pr}^* and the approximated value $\lambda_1 R_{\mathrm{Pr}}^1 + \lambda_2 R_{\mathrm{Pr}}^2 + \lambda_3 R_{\mathrm{Pr}}^3$ at (U^*, α^*) can be expressed as

$$\begin{aligned} |\Phi| &= |\lambda_1 R_{\mathrm{Pr}}^1 + \lambda_2 R_{\mathrm{Pr}}^2 + \lambda_3 R_{\mathrm{Pr}}^3 - R_{\mathrm{Pr}}^*| \\ &= |R_{\mathrm{Pr}}^1 + (R_{\mathrm{Pr}}^2 - R_{\mathrm{Pr}}^1)\lambda_2 + (R_{\mathrm{Pr}}^3 - R_{\mathrm{Pr}}^1)\lambda_3 - R_{\mathrm{Pr}}^*| \\ &= \left| R_{\mathrm{Pr}}^1 + (R_{\mathrm{Pr}}^2 - R_{\mathrm{Pr}}^1)\frac{U^* - U_1}{\Delta U} + (R_{\mathrm{Pr}}^3 - R_{\mathrm{Pr}}^2)\frac{\alpha^* - \alpha_1}{\Delta\alpha} - R_{\mathrm{Pr}}^* \right|. \end{aligned}$$

Hence, we have

$$\frac{\partial\, |\Phi|}{\partial \Delta U} = -\frac{|(U^* - U_1)(R_{\mathrm{Pr}}^2 - R_{\mathrm{Pr}}^1)|}{\Delta U^2} < 0 \text{ and } \frac{\partial\, |\Phi|}{\partial \Delta\alpha} = -\frac{|(\alpha^* - \alpha_1)(R_{\mathrm{Pr}}^3 - R_{\mathrm{Pr}}^2)|}{\Delta\alpha^2} < 0$$

which means that the distance decreases as $\Delta U \to 0$ and $\Delta\alpha \to 0$. With $\Delta U \to 0$ and $\Delta\alpha \to 0$, the three vertices are close enough, and thus we have $U^* \to U_1$ and $\alpha^* \to \alpha_1$ and then $|\Phi| \to 0$. Because $2N \to \infty$ and $2M \to \infty$ result in $\Delta U \to 0$ and $\Delta\alpha \to 0$, the proof is complete. □

Using the model transformation in Section 3.1.1, the independent variables are the same in Eqs. (24) and (25). Therefore, the linearization scheme for both equations can be identical. In this way, the numbers of variables and constraints can greatly decrease, leading to computational efficiency. Hence, we have the following set of mixed-integer constraints that linearize Eq. (25), which is (27), (29)–(33), and

$$\sum_{v \in V(T)} \lambda_v \cdot R_O^{w,p}(v_x, v_y) = \widehat{R}_O^{w,p} \tag{34}$$

where $\widehat{R}_O^{w,p}$ is the approximated value of $R_O^{w,p}$.

Define vectors $\boldsymbol{\gamma} = [\gamma_k]$, $\boldsymbol{\lambda} = [\lambda_v]$, $\widehat{\mathbf{R}}_{\mathbf{O}} = [\widehat{R}_O^{w,p}]$, and $\widehat{\mathbf{R}}_{\mathbf{Pr}} = [\widehat{R}_{\mathrm{Pr}}^{w,p}]$. The original problem (P0) can then be approximated by the following:

(P3)

$$\min_{\mathbf{y},\mathbf{U},\boldsymbol{\alpha},\widehat{\mathbf{R}}_{\mathbf{O}},\widehat{\mathbf{R}}_{\mathbf{Pr}},\boldsymbol{\lambda},\boldsymbol{\gamma}} Z_3 = -\sum_{w \in W} d^w \sum_{p \in P^w} \widehat{R}_O^{w,p} \tag{35}$$

subject to

$$\sum_{a\in\widetilde{A}} y_a l_a \tau \leqslant B \tag{36}$$

$$y_a \in \{0, 1\} \quad \forall\, a \in \widetilde{A} \tag{37}$$

$$\sum_{p\in P^w} \widehat{R}_{\mathrm{Pr}}^{w,p} = 1 \quad \forall\, w \in W \tag{38}$$

$$U_p^w = U_{p,w}^0 + \varphi \sum_{a\in A} \frac{l_a}{L_p} y_a \delta_{a,p}^w \quad \forall\, p \in P^w,\, w \in W \tag{39}$$

$$\sum_{v\in V(T)} \lambda_v \cdot (v_x,\, v_y) = (U_p^w,\, \alpha^w) \tag{40}$$

$$\sum_{v\in V(T)} \lambda_v \cdot R_{\mathrm{Pr}}^{w,p}(v_x,\, v_y) = \widehat{R}_{\mathrm{Pr}}^{w,p} \tag{41}$$

$$\sum_{v\in V(T)} \lambda_v \cdot R_O^{w,p}(v_x,\, v_y) = \widehat{R}_O^{w,p} \tag{42}$$

$$\lambda_v \geqslant 0 \quad \forall\, v \in V(T) \tag{43}$$

$$\sum_{v\in V(T)} \lambda_v = 1 \tag{44}$$

$$\sum_{v\in L(k)} \lambda_v \leqslant \gamma_k \tag{45}$$

$$\sum_{v\in R(k)} \lambda_v \leqslant 1 - \gamma_k, \text{and} \tag{46}$$

$$\gamma_k \in \{0, 1\} \quad \forall\, k \in K \tag{47}$$

where

$$PS_p^w = \sum_{a\in A^p} \frac{l_a}{L_p} \frac{1}{\sum_{q\in P^w} \delta_{a,q}^w} \quad \forall\, p \in P^w,\, w \in W \tag{48}$$

One advantage of the reformulated problem (P3) is that it is formulated as an MILP, which can be solved with traditional algorithms for MILPs, such as the branch and bound algorithm, or by many commercial solvers, such as CPLEX, Gurobi, and Mosek, to its global optimum, which is also a nearly global optimum to the original problem (P0). For any obtained **y**, the real objective function value can be recalculated by solving a PSL traffic assignment problem.

The following proposition, Proposition 3, proves that problem (P3) approaches the original problem (P0) when the area of each polytope tends to zero.

Proposition 3. *For the given feasible regions* $[U_p^{w,\min}, U_p^{w,\max}]$ *and* $[\alpha^{w,\min}, \alpha^{w,\max}]$, *when the numbers of intervals* $2N \to \infty$ *and* $2M \to \infty$, *problem* (*P*3) → *the original problem* (*P*0).

Proof. This result can be obtained via Propositions 1 and 2. □

Remark.. According to Proposition 3, the global optimality of the solution determined by this approach is subject to the linearization scheme adopted. The finer the linearization scheme, the more precise the solution. According to the numerical experiments, the linearized model requires few breakpoints to guarantee solution quality. Generally, five to eleven breakpoints for each region of the variable, even for a large network such as the Anaheim network, is sufficient to obtain an accurate solution.

3.2. Proposed matheuristic

Generally, although global optimization methods can guarantee an exact solution to the linearized problem, the computational efficiency of most global optimization methods is not satisfactory, especially for large instances. In this subsection, a matheuristic is presented to greatly improve the computational efficiency for solution of large instances while guaranteeing solution quality. In particular, a surrogate-model–based solution algorithm is first used to update the feasible region of the problem, and the global optimization method presented in Section 3.1 is then applied to solve the model for problem (P3) within the updated feasible region, which likely contains an optimum to the original problem. In this way, only a small MILP must be solved with the global optimization method, thus improving the computational efficiency. Note that relative to direct use of the surrogate-model–based algorithm to solve the problem, our proposed matheuristic can maintain the algorithm's computational efficiency while improving the solution quality. Fig. 2 shows the flowchart of the matheuristic.

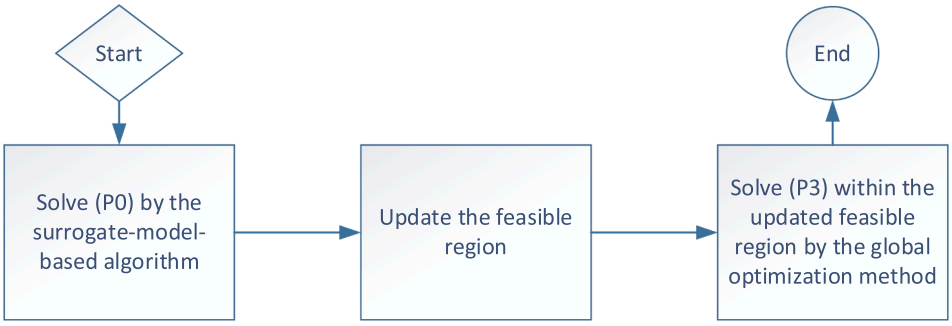


Fig. 2. Flowchart of the matheuristic.

3.2.1. Surrogate-model–based solution algorithm

A surrogate-model–based solution algorithm is a type of efficient heuristic that is usually used to solve complicated nonlinear problems for which derivative information can hardly be obtained. The core of the surrogate-model–based solution algorithm is to construct and iteratively update a simple surrogate or response surface model to approximate the original complicated problem. Solving the original problem is then converted into estimation of solutions with the surrogate model. Exactly one point is selected with the surrogate model in each iteration and evaluated in the original model for problem (P0). This characteristic generally allows the surrogate-model–based solution algorithm to outperform other heuristics, such as genetic algorithm, in computational efficiency. Details regarding surrogate-model–based solution algorithms were published by Regis and Shoemaker (2007) and Regis (2011).

The key decision for this problem is to determine the bike path construction plan, which is encoded by the binary vector $\mathbf{y} = [y_a]_{a\in\widetilde{A}}$. Given a known vector $\mathbf{y}$, the objective function value of the original model (i.e., social welfare) can be calculated. Hence, the surrogate model $S(\mathbf{y})$ in this problem is constructed in terms of an independent binary vector $\mathbf{y}$ to approximate the objective function of (P0). The stochastic surrogate-model–based algorithm described by Liu and Wang (2017) that was used to solve problems with discrete variables is extended to solve the bike network design problem. The detailed procedure of the stochastic surrogate-model–based algorithm is stated below:

Step 1: Initialization.

Step 1.1: Find a set of initial starting points $I_0 = \{\mathbf{y}_1, \mathbf{y}_2, ...\mathbf{y}_{n_0}\}$ that contains at least one feasible vector $\mathbf{y}_i$ in I_0, where the subscript i is the index of $\mathbf{y}$ in I_0, $n_0 = dim + 1$, and dim is the dimension of $\mathbf{y}$. For each initial starting point, evaluate this vector using the original model and obtain the objective function value, that is, $Z(\mathbf{y}_i)$. Find the best feasible solution $\mathbf{y}_{best}$ among the set I_0. Set the iteration number $n = n_0$ and the set of evaluated points $I_n = I_0$.

Step 1.2: Initialize the probability of perturbing a dimension of the current best feasible solution for generation of random candidate points in Step 2, denoted by p_{slct}. Initialize the counters for consecutive successes and failures: $C_{succ} = 0$ and $C_{fail} = 0$. Define a positive integer κ and a series of weights μ_i for the response surface score used in Step 2.3 that satisfy the condition $0 \leqslant \mu_1 \leqslant ...\leqslant\mu_\kappa \leqslant 1$. Define a positive integer parameter $C_{fail}^{\max}$ as the maximum number of consecutive failures and a positive integer parameter $C_{succ}^{\max}$ as the maximum number of consecutive successes.

Step 2: Iteration. *While* the termination condition is not satisfied, do:

Step 2.1: Construct $S_1(\mathbf{y})$ in iteration 1 or update $S_n(\mathbf{y})$ in iteration n when $n \geqslant 2$, where $S_n(\mathbf{y})$ is the surrogate model in the nth iteration, for approximation of the original bike path network design model's objective function value $Z(\mathbf{y})$ by using evaluated candidate points in $B_n = \{(\mathbf{y}_i, Z(\mathbf{y}_i)), \mathbf{y}_i \in I_n\}$.

Step 2.2: Randomly generate a set of t candidate feasible vectors $E_n = \{\mathbf{q}_1, ...,\mathbf{q}_t\}$ to estimate the true objective function value with the surrogate model $S_n(\mathbf{y})$, where the dimension of $\mathbf{q}_j$ is dim and t can be unequal to dim. *For* each $\mathbf{q}_j$, $j \in \{1, ...,t\}$,

Step 2.2.1: Randomly generate dim uniformly distributed numbers $u_1,..., u_{dim}$ in [0,1]. Select the index i of u_i from the set $I_{pert} = \{i|\ u_i < p_{slct}, i \in [1, dim]\}$, where p_{slct} is the probability of perturbing a dimension of the current best feasible solution for generation of random candidate points. If $I_{pert} = \varnothing$, uniformly and randomly select an index i from the set $\{1,..., dim\}$ and let $I_{pert} = \{i\}$. For each index $i \in I_{pert}$, $\mathbf{q}_j^{(i)} = 1 - \mathbf{y}_{best}^{(i)}$, where $\mathbf{q}_j^{(i)}$ is the ith element of $\mathbf{q}_j$ and a similar meaning holds for $\mathbf{q}_{best}^{(i)}$; for index $i \notin I_{pert}$, $\mathbf{q}_j^{(i)} = \mathbf{y}_{best}^{(i)}$.

Step 2.2.2: Calculate the total construction cost for $\mathbf{q}_j$. If $\mathbf{q}_j$ fails to satisfy the budget constraint, randomly select an index $i \in \{1, ...,dim\}$, where $\mathbf{q}_j^{(i)} = 1$ and let $\mathbf{q}_j^{(i)} = 0$. Calculate the total construction cost and check the budget constraint again. Repeat this process until $\mathbf{q}_j$ satisfies the budget condition. *End* this *for* iteration.

Step 2.3: Apply the surrogate model $S_n(\mathbf{y})$ to select the candidate solution to be evaluated in the space of $Z(\mathbf{y})$ from the generated t candidate feasible points.

Step 2.3.1: Calculate $S_n(\mathbf{q})$ for each $\mathbf{q} \in E_n$, $S^{\min} = min\{S_n(\mathbf{q}), \mathbf{q} \in E_n\}$ and $S^{\max} = \max\{S_n(\mathbf{q}), \mathbf{q} \in E_n\}$. Compute the response surface score $Q_n^S(\mathbf{q})$ for each $\mathbf{q} \in E_n$: if $S^{\max} \neq S^{\min}$, $Q_n^S(\mathbf{q}) = (S_n(\mathbf{q}) - S^{\min})/(S^{\max} - S^{\min})$; otherwise $Q_n^S(\mathbf{q}) = 1$.

Step 2.3.2: Calculate the minimum distance from previously evaluated points. For each $\mathbf{q} \in E_n$, this distance is defined as $D_n(\mathbf{q}) = min_{1\leqslant i\leqslant n}\ \|\mathbf{q} - \mathbf{y}_i\|$, $\mathbf{y}_i \in I_n$. The symbol $\|\cdot\|$ stands for the Euclidean norm. Calculate $D^{\min} = min\{D_n(\mathbf{q}), \mathbf{q} \in E_n\}$ and $D^{\max} = \max\{D_n(\mathbf{q}), \mathbf{q} \in E_n\}$ and compute the distance score: for each $\mathbf{q} \in E_n$, the score

$Q_n^D(\mathbf{q}) = (D^{\max} - D_n(\mathbf{q}))/(D^{\max} - D^{\min})$ if $D^{\max} \neq D^{\min}$; otherwise $Q_n^D(\mathbf{q}) = 1$.

Step 2.3.3: Set the weight for the response surface score $\omega_n^S = \begin{cases} \mu_{\bmod(n-n_0,\kappa)} & \text{if } \bmod(n-n_0,\kappa) \neq 0 \\ \mu_\kappa & \text{otherwise} \end{cases}$ and the weight for the distance score $\omega_n^D = 1 - \omega_n^S$. Compute the final weighted score $Q_n(\mathbf{q}) = \omega_n^S Q_n^S(\mathbf{q}) + \omega_n^D Q_n^D(\mathbf{q})$ for each $\mathbf{q} \in E_n$ and select the solution $\mathbf{q}^*$ with the minimum weighted score. Let $\mathbf{y}_{n+1} = \mathbf{q}^*$.

Step 2.4: Evaluate $\mathbf{y}_{n+1}$ using the model for (P0).

Step 2.5: Update the current best feasible solution $\mathbf{y}_{best} = \mathbf{y}_{n+1}$ and set $C_{succ} = C_{succ} + 1$ and $C_{fail} = 0$ if $\mathbf{y}_{n+1}$ is feasible and $Z(\mathbf{y}_{n+1}) < Z(\mathbf{y}_{best})$; otherwise, set $C_{fail} = C_{fail} + 1$ and $C_{succ} = 0$.

Step 2.6: Set $p_{slct} = 0.5p_{slct}$ and $C_{fail} = 0$ if $C_{fail} \geqslant C_{fail}^{\max}$. If $C_{succ} \geqslant C_{succ}^{\max}$, set $p_{slct} = 2p_{slct}$ and $C_{succ} = 0$.

Step 2.7: Update the set of evaluated points $I_{n+1} = I_n \cup \{\mathbf{y}_{n+1}\}$ and the iteration number $n = n + 1$. *End* the *while* iteration.

Step 3: Output. Return the current best feasible solution $\mathbf{y}_{best}$. □

Remarks.

1. In Step 1.1, at least one feasible point $\mathbf{y}_i$ and a set of initial starting points $I_0 = \{\mathbf{y}_1, \mathbf{y}_2, ...\mathbf{y}_{n_0}\}$ are needed to construct the initial surrogate model. For purposes of illustration, without loss of generality, we suppose that $\mathbf{y}_1$ is a feasible point. To obtain a feasible $\mathbf{y}_1$, a random point was generated and then processed via Step 2.2.2 to make it satisfy the budget constraint. Note that the initial points should be affinely independent to fit the initial surrogate model, and the obtained $\mathbf{y}_1$ was applied to generate all other starting points. Define $\{\mathbf{e}_1, ...,\mathbf{e}_{\dim}\}$ as the natural basis of $\mathbb{R}^{dim}$. The other starting points $\{\mathbf{y}_2, ...\mathbf{y}_{n_0}\}$ can be defined as $\{|\mathbf{y}_1 - \mathbf{e}_1|, ...,|\mathbf{y}_{n_0} - \mathbf{e}_{\dim}|\}$. This is only one viable method to generate an initial point set; other methods can also be proposed and applied. Also, please note that infeasible initial points can be used when fitting the surrogate model in Step 2.1.
2. In Step 2.3, two criteria are considered in the selection of evaluation points: the value estimated with the surrogate model and the minimum distance from the previously evaluated points. The inclusion of the second criterion can promote the global search of the feasible region and improve the construction and update of the surrogate model.
3. Unlike solving continuous problems, solving the discrete bike path network design problem involves the use of the perturbing probability p_{slct}. The main function of this probability is to adjust the average number of elements in $\mathbf{y}_{best}$ that undergo perturbation, that is, to control the search area for candidate solutions. Generally, p_{slct} is enlarged if the algorithm encounters a number of consecutive successes; p_{slct} is reduced to search in a small neighborhood if the algorithm encounters a number of consecutive failures. This rule ensures that the search area decreases over iterations after the global optimum is determined and that the whole algorithm is convergent. The stopping criterion can be set as $p_{slct} \cdot dim < 1$. That is, if p_{slct} is sufficiently small, the random points generated in Step 2.2 will be no different from the current best solution, and the algorithm will stop.

The above surrogate-model–based solution method is an efficient heuristic. It can be applied directly to solve large bike path network design problems, just as Liu and Wang (2017) used it to solve a location problem. However, the quality of $\mathbf{y}_{best}$ obtained directly from the heuristic cannot be guaranteed. Unlike Liu and Wang (2017), we do not only rely upon the single solution $\mathbf{y}_{best}$ that is directly obtained from the algorithm. In the next subsection, the set of evaluated points I_n in the last iteration of the surrogate-model–based algorithm will be used to update the feasible region of problem (P3).

3.2.2. Novel hybrid of a surrogate-model–based heuristic and the exact method

Combination of heuristic and global optimization methods is of great significance. The proposed novel solution method possess the merits of both, that is, computational efficiency and high solution quality. The main idea is to reasonably reduce the size of problem (P3) with the set of evaluated points I_n obtained from the surrogate-model–based algorithm. At least, the novel method finds a solution no worse than that determined by the surrogate-model–based algorithm.

The detailed steps of the novel solution method are given as follows:

Step 1: Solve the original problem (P0) with the surrogate-model–based algorithm to obtain a set of evaluated points I_n.

Step 2: Remove duplicate points and select the best χ evaluated points $I' = \{\mathbf{y}_1, \mathbf{y}_2, ...\mathbf{y}_\chi\}$ from I_n. Substitute each $\mathbf{y}_i \in I'$ into the original problem (P0) and calculate $U_{p,i}^w$ and α_i^w for each route and OD pair. Set the new upper and lower bounds for U_p^w as $U_p^{w,\max} = \max\{U_{p,i}^w, i = 1. .\chi\}$ and $U_p^{w,\min} = \min\{U_{p,i}^w, i = 1. .\chi\}$, respectively. Correspondingly, set the new upper and lower bounds for α^w as $\alpha^{w,\max} = \max\{\alpha_i^w, i = 1. .\chi\}$ and $\alpha^{w,\min} = \min\{\alpha_i^w, i = 1. .\chi\}$, respectively.

Step 3: Formulate and solve the MILP for problem (P3) upon the updated feasible region with the global optimization method in Section 3.1 and obtain the final solution. □

Remarks. The value of χ has a great effect on the final solution. In an extreme condition (i.e., $\chi = 1$), the final solution is exactly the same as the solution obtained from the surrogate-model–based algorithm. Thus, the novel solution method is no worse than the pure surrogate-model–based algorithm. A larger χ leads to a larger feasible region for the approximated problem (P3), which may result in a longer computational time, but it also carries a greater likelihood of obtaining a better solution. Therefore, we should carefully choose χ to make a tradeoff between computational efficiency and solution quality.

4. Numerical studies

Numerical experiments were conducted to analyze the characteristics of the developed bicycle network design model and to

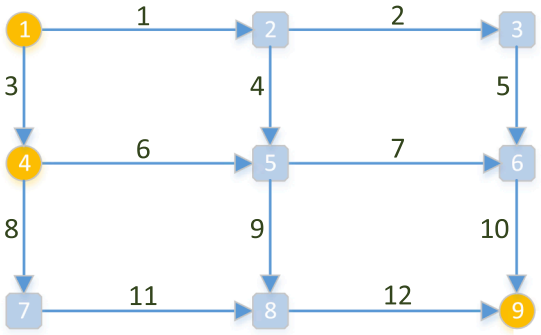


Fig. 3. Nine-node network for the numerical experiments of Example 1.

evaluate the performance of the proposed solution methods. Three networks are adopted in this section: a nine-node network, the widely used Sioux Falls network, and the Anaheim network. The small nine-node network is used to show (1) the accuracy of the proposed global optimization method; (2) the effects of various investment budgets; (3) the benefit inequity issue between various OD pairs; and (4) the differences in the results between our model and that of a bicycle network design model from the literature. The performance of the proposed global optimization method and the novel solution method on larger network applications was tested with the other two example networks.

All experiments were run on a PC with Windows 10, a 4.2-GHz Intel Core i7-7700 k processor, and 16 GB of RAM. The academic version of the commercial solver CPLEX was used as an external solver of YALMIP (Löfberg, 2004) in MATLAB R2014b to solve all MILPs in the experiments. Unless stated otherwise, the parameter φ in the utility function was set to 1.57 (Broach et al., 2012) for all experiments.

4.1. Example 1: The nine-node network

The nine-node test network shown in Fig. 3 was first adopted for the model and solution method analysis. We assume that each link (i.e., roadway) is a candidate for bike path construction. The length of each link is given in Table 1. Let the construction cost for any bike path be 2 units per mile. Two OD pairs are considered in most experiments, which are OD pair 1- > 9 with the demand of 10 units and OD pair 4- > 9 with the demand of 20 units. All existing routes between each OD pair are considered in this section. Table 2 provides the current utility of each route.

4.1.1. Performance of the proposed global optimization method

In this experiment, the adopted linearization scheme of the proposed method uses only five breakpoints, including the endpoints, to split the feasible regions of U_p^w and α^w, which should not be treated as a very fine linearization scheme.

Table 3 compares the results obtained with the proposed method and those obtained with enumeration. The Z_a column gives the optimal objective function value of the MILP for problem (P3), and the Z_r column gives the real objective function value of the network by evaluation of the obtained construction plan from the MILP via the PSL traffic assignment problem. In this specific case, Z_r also equals the optimal objective function value obtained by enumeration.

Table 3 shows that the *Gap* between Z_a and Z_r is relatively small. The maximum gap is 0.20%, which occurs when a given budget is 2 units; the smallest gap is nearly zero, which occurs when the budget equals 0.5 and 6.5 units. Moreover, we compare the solutions obtained with the proposed method with those obtained with enumeration. The results show that the global optima obtained with the proposed method are the same as obtained with the enumeration method. This implies that a relatively small *Gap* exists between the objective function value obtained directly from the MILP and the real objective function value obtained by evaluating the optimal solution of the MILP by solving the PSL traffic assignment problem.

To illustrate the accuracy of the MILP for (P3), we compare the values of $\widehat{R}_{Pr}^{w,p}$ and $\widehat{R}_{O}^{w,p}$ obtained from the MILP for (P3) based on the linear combination of three vertex values and the corresponding real values (i.e., $R_{Pr}^{w,p}$ and $R_{O}^{w,p}$) obtained from the PSL traffic assignment. The budget was set to 5 units. Fig. 4 shows the corresponding optimal bike path construction plan. The dashed lines represent bike paths. The comparison results are given in Fig. 5. The dark colored bars represent the real values of $R_{Pr}^{w,p}$ and $R_{O}^{w,p}$ for all routes between each OD pair, and OD1.2 refers to route 2 of OD pair 1, etc. The light-colored bars represent the values of $\widehat{R}_{Pr}^{w,p}$ and $\widehat{R}_{O}^{w,p}$ (the approximate values of $R_{Pr}^{w,p}$ and $R_{O}^{w,p}$, respectively). The approximate values obtained with the linearization method are very close to the corresponding real values. In fact, the maximal gap between $\widehat{R}_{Pr}^{w,p}$ and $R_{Pr,r}^{w,p}$ is about 1.7%, whereas the maximal gap is

Table 1
Length of each link.

Link	1	2	3	4	5	6	7	8	9	10	11	12
Length (mile)	0.6	0.5	0.3	0.8	0.6	0.3	0.4	0.5	0.8	0.6	0.3	0.5

Table 2
Utility of each route before introduction of bicycle paths.

OD pair	Route	Utility	OD pair	Route	Utility
1- > 9	[1,2,5,10]	−7.5	1- > 9	[3,8,11,12]	−6
	[1,4,9,12]	−9.5			
	[1,4,7,10]	−8.7	4- > 9	[6,7,10]	−5.8
	[3,6,7,10]	−7.7		[6,9,12]	−6.8
	[3,6,9,12]	−8.7		[8,11,12]	−6.3

Table 3
Comparison of results between the proposed global optimization method and enumeration.

Budget	Z_a	Z_r	Gap
0.5	187.9972	187.9972	0.00%
2	163.8110	164.1422	0.20%
3.5	151.0849	151.1211	0.02%
5	145.8699	145.6688	0.14%
6.5	139.5155	139.5147	0.00%

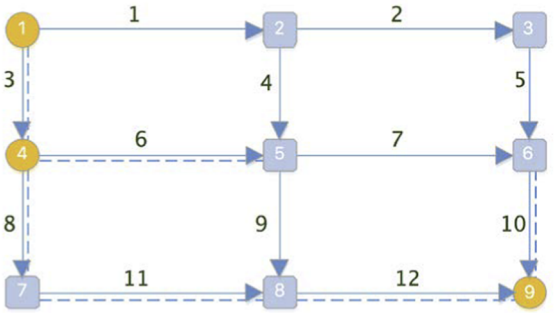
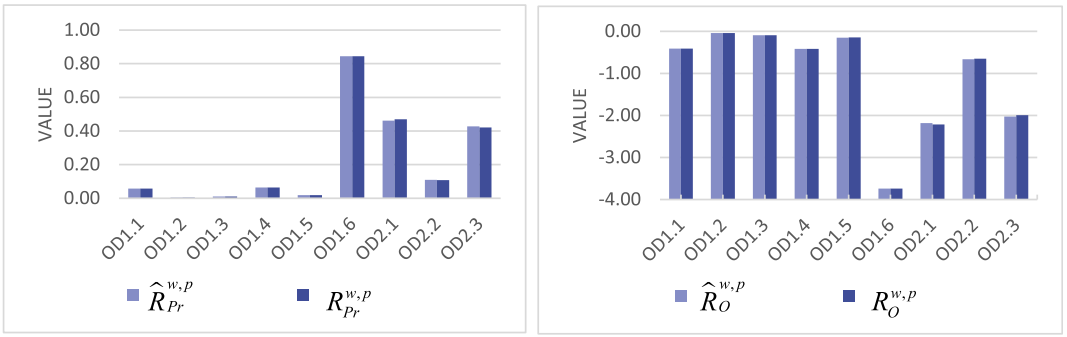


Fig. 4. Bike path construction plan at a budget of 5 units.



(a) $\hat{R}_{Pr}^{w,p}$ *and* $R_{Pr}^{w,p}$ (b) $\hat{R}_{O}^{w,p}$ *and* $R_{O}^{w,p}$.

Fig. 5. Comparisons between the values of (a) $\hat{R}_{Pr}^{w,p}$ *and* $R_{Pr}^{w,p}$ and (b) $\hat{R}_{O}^{w,p}$ *and* $R_{O}^{w,p}$ at a budget of 5 units.

1.5% for $\hat{R}_{O}^{w,p}$ and $R_{O}^{w,p}$. This implies that the proposed MILP is a good approximation for the original nonlinear programming model.

To analyze the effect of the value of the parameter φ on the optimal construction plan and the objective function value, we set $\varphi = 1.57\mu$, where μ is the multiplier to the reference value of 1.57 obtained from the literature, and varied the multiplier instead of changing the value of φ directly. Table 4 shows the effect of the multiplier on the optimal construction plan. The optimal construction plan differs only when the multiplier is extremely small or large (e.g., when $\mu = 0.005$ or 2); otherwise, the optimal construction plan remains the same (e.g., $\mu \in [0.05, 1.5]$). This table also shows that the number of links in the optimal construction plan does not necessarily increase with the multiplier. For example, at $\mu = 1.5$, six bike paths are built, but only four bike paths are constructed at $\mu = 2$. This observation implies that the number of links in the optimal construction plan does not necessarily increase with the value

Table 4
Impact of the value of the multiplier on the optimal construction plan.

Multiplier μ	Optimal plan
0.005	(6,7,8,10,12)
0.05	(3,6,8,10,11,12)
0.5	(3,6,8,10,11,12)
1	(3,6,8,10,11,12)
1.5	(3,6,8,10,11,12)
2	(3,6,11,12)

of φ. Fig. 6 shows that the objective function value gradually decreases as the value of μ increases, which is consistent with our understanding. A larger value of μ implies a larger value of φ, which means that the effect of bike paths on the route utility is larger. If the proportion of bike paths on a route is fixed, a larger value of φ makes the route utility less negative or the absolute value of the route utility smaller. If the proportions of the bike paths of all routes are fixed (i.e., the optimal plan is fixed), the absolute value of the total route utility of cyclists (i.e., the objective function value) decreases as φ increases, as reflected in the interval of μ between 0.05 and 1.5. Even if the optimal construction plan changes as the value of φ increases from a large value to a very large value (e.g., the corresponding value of μ increases from 1.5 to 2), the optimal objective function value cannot increase because the optimal construction plan at a large value of φ only becomes feasible with a large value of φ.

Table 5 shows the effect of the number of breakpoints on the objective function values Z_a and Z_r and the computational time of the global optimization method. In this small example, even with five breakpoints including the endpoints for each variable, the solution obtained is globally optimal compared with that of the enumeration method. The gap between Z_a and Z_r is quite small and generally decreases from 0.2018% to 0.0647% as the number of breakpoints increases. However, the pursuit of a more precise solution requires a longer computation time. This is also one reason why the second solution method is presented for large network instances; even the first solution method can guarantee that a global optimum will be obtained with a certain degree of accuracy.

4.1.2. Effects of the investment budget

To illustrate the effects of the investment budget, Table 6 contrasts changes in the optimal construction plan with gradual increases in the budget. Three observations can be made. First, with a very small initial budget, say 0.5 unit, when the budget increases, the number of links with bike paths generally increases. In particular, the number of links with bike paths increases from two to seven when the budget increases from 2 to 6.5 units. Second, with a very large initial budget, the optimal plan does not change when the budget is further increased. In this example, the optimal plan remains the same when the budget increases from 6.5 to 8 units. This serves as a good reminder to the transport planning department that an upper bound for construction investment indeed exists, and exceeding this limit does not achieve better performance for the whole network. Third, in this example, we find that the links in the optimal plan with a lower budget also appear in the optimal plan with a larger budget. This conclusion may be true for this specific example, but it may not be true for general cases. Indeed, Section 4.1.3 gives an example that shows this possibility.

Table 7 shows a Braess-like paradox in which another bike path is built on link 9 in the optimal plan obtained for a budget of 8 units. The resultant objective function value does not decrease relative to the objective function value of the optimal plan but rather increases from 139.51 to 139.91 (i.e., the total route utilities of all cyclists decrease). However, the route utilities of cyclists between the two OD pairs are not reduced and are improved in some cases. The total utilities decrease because a greater portion of the bicycle flow chooses a route other than that with the highest utility according to the cyclists' route choice behavior or the PSL route choice principle.

Fig. 7 shows the changes in the link flow in the network with gradual increases in the budget. In this figure, solid lines are used to represent the links in the optimal construction plan with a budget of 6.5 units, which we call type 1 links. Dotted and dashed lines are used to represent the links that do not appear in the optimal plan, which we call type 2 links. Please note that the lines between two markers do not represent the real flow on the link; rather, they are drawn to provide the reader with an intuitive sense of the changing

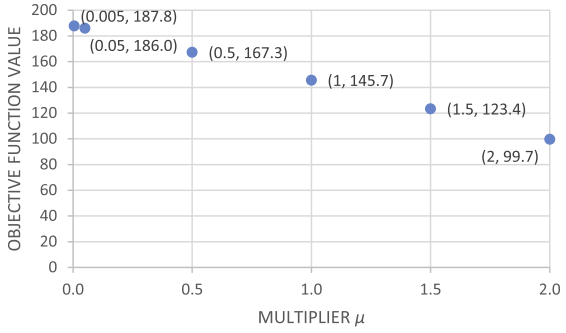


Fig. 6. Impact of the value of the multiplier on the objective function value.

Table 5
Effect of the number of breakpoints on objective function values and computation time.

No. of breakpoints	Z_a	Z_r	Gap	Time (s)
5	163.8110	164.1422	0.2018%	0.1385
9	163.9818	164.1422	0.0977%	0.2079
13	164.0360	164.1422	0.0647%	0.3518

Table 6
Bike path construction plan under various budgets.

Budget	Optimal plan
0.5	–
2	(8,12)
3.5	(3,8,11,12)
5	(3,6,8,10,11,12)
6.5	(3,6,7,8,10,11,12)
8	(3,6,7,8,10,11,12)

Table 7
Comparison between the best plan and another plan that incorporates the optimal plan when both budgets equal 8 units.

OD pair	Route no.	Optimal plan	(3,6,7,8,10,11,12)	Another plan	(3,6,7,8,***9***,10,11,12)
		Z	139.51	Z	139.91
		Probability	Route utility	Probability	Route utility
1- > 9	1	0.06	−7.09	0.05	−7.09
	2	0.00	−9.21	0.01	−8.74
	3	0.01	−8.05	0.01	−8.05
	4	0.09	−6.13	0.09	−6.13
	5	0.02	−7.79	0.03	−7.13
	6	0.82	−4.43	0.80	−4.43
4- > 9	1	0.59	−4.23	0.54	−4.23
	2	0.08	−6.02	0.17	−5.23
	3	0.33	−4.73	0.30	−4.73

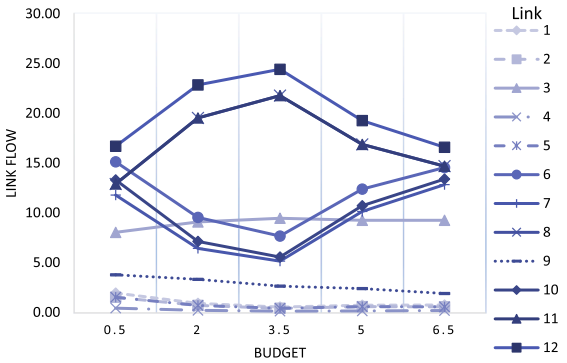


Fig. 7. Bicycle flow on each link with various budgets.

trend of link flow. We observe that the cycling flow on type 2 links is generally much smaller than that on type 1 links. Moreover, the changes of flow on these links are less evident, but a slight downward trend can be seen. Type 1 links can be separated into two groups; in one group, which includes links 8, 11, and 12, the link flow first increases and then decreases, and in the other group, which includes links 6 and 10, shows an opposite trend. This finding illustrates that when a bike path network is constructed, a portion of the cycling flow on routes without bike paths are attracted to the routes with new bike paths, which coincides with the real-life situation. In addition, as construction investment continues, some of the attracted cycling flow returns to its original route.

4.1.3. Benefit inequity between various OD pairs

In this example, an additional OD pair 1- > 3 with a demand level of 10 was added to the test network to explain the inequity issue between various OD pairs. Table 8 shows the utility of each OD pair in various budget scenarios and the beneficial OD pairs after

Table 8
Beneficial OD pairs with various investment budgets.

Budget	Optimal plan	OD pair utility			Beneficial OD pairs
		1- > 9	4- > 9	1- > 3	
0	–	−65.69	−122.31	−60.00	–
0.5	–	−65.69	−122.31	−60.00	–
2	(8,12)	−54.40	−109.74	−60.00	1- > 9,4- > 9
3.5	(3,8,11,12)	−47.53	−103.59	−60.00	1- > 9,4- > 9
5	(1,2,8,11,12)	−51.59	−103.59	−44.30	1- > 9,4- > 9,1- > 3
6.5	(1,2,6,7,8,10,11)	−57.44	−92.26	−44.30	1- > 9,4- > 9,1- > 3
8	(1,2,3,6,7,8,10,11,12)	−49.30	−90.86	−44.30	1- > 9,4- > 9,1- > 3

implementation of the optimal plan of the bicycle network in each situation. Fig. 8 compares the utility of each OD pair in various budget scenarios with the corresponding utility of the original network (i.e., with a budget of 0 units). We note that in some cases, such as with budgets of 2 and 3.5 units, not all OD pairs benefit from the new bicycle facility construction. Some OD pairs receive a greater increase in utility than the others. For example, in the case with a budget of 2, OD pair 1- > 9 receives a utility improvement of 17.2%, whereas the utility of OD pair 1- > 3 remains unchanged. In another aspect, we take a system level perspective in the design of the bike path network. Therefore, the OD pair with the least utility may not benefit from improvement of the budget—for example, the utility of OD pair 4- > 9 remains unchanged when the budget is increased from 3.5 to 5 units. Moreover, the utility for OD pair 1- > 9 even decreases when the budget is increased from 3.5 to 6.5 units. This phenomenon is also normal due to the system optimal objective of the network design model.

4.1.4. Comparison with a bicycle network design model from the literature

This subsection compares the results obtained with our model with those obtained with the model presented by Duthie and Unnikrishnan (2014). The test bed remains the network shown in Fig. 3 but with three OD pairs. Two OD pairs are the same as those in Section 4.1.1, and the new OD pair is OD pair 2- > 9, with a demand of 10. Although the model of Duthie and Unnikrishnan (2014) does not require the demand of each OD pair, we still provide them for comparison with our model. Other parameters remain the same as the values set in Section 4.1.1. In fact, Duthie and Unnikrishnan's model mainly considers the route length constraint, and its objective is to minimize the construction investment. Their design ensures that at least one route is suitable for cycling between each OD pair and that the length of each route of each OD pair is within a given tolerance and is proportional to the shortest route distance and smaller than the maximum length that is acceptable to a cyclist between each OD pair. Their model does not consider the actual route utility (which includes the effects of roadway slope and other bicycle facilities) or the cyclists' route choice behavior (which is captured by the PSL traffic assignment in our model).

The bike path construction solutions for the test network can be obtained separately from the two models with the same input data. In the proposed model, the investment budget is an input parameter, but in the model from the literature, it is an output. Moreover, the model from the literature requires an input of the maximum route length acceptable to cyclists. Therefore, the literature model was first solved by assuming that the maximum route length is 20% longer than the shortest route between each OD pair. For a fair comparison with our model, the intersection construction cost (which is required in the literature model) was not included in the investment budget. We then used the obtained minimum investment budget to solve the proposed model. Table 9 compares the optimal plans obtained from the literature model and from the proposed model, both with respect to the investment budget (i.e., the objective function value of the literature model) and to the total utility (i.e., the objective function value of the proposed model). Table 9 shows that the optimal plans obtained from the two models differ considerably, which is understandable because the planning principles are totally different — one considers the route distance only, and the other considers the actual route utility. Because our proposed model considers cyclists' realistic route choice behavior, it shows better system performance than the literature model, with a 2.5% increase in total utility for this small example. Moreover, the optimal construction plan obtained by our

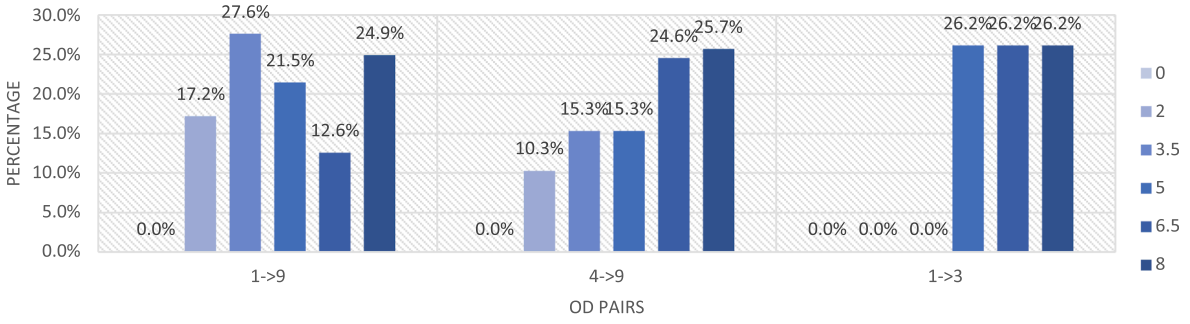


Fig. 8. Percentage increases in OD pair utilities with various budgets.

Table 9
Optimal plans obtained from the literature model and from our model.

Model	Optimal plan	Budget	Total utility of all cyclists
Duthie and Unnikrishnan (2014)	(3,4,6,7,10)	4.8	−210.154
Our proposed model	(3,6,7,10,11,12)	4.6	−205.016

model requires a lower investment budget. In summary, the plans obtained from the two models can differ greatly, and it is important to capture cyclists' realistic route choice behavior from the aspects of total utility maximization and the budget.

The literature model's computation time is about 0.02 s, whereas that of our proposed model is about 0.30 s because the proposed model considers cyclists' route choice behavior and solving the traffic assignment requires more computation time than solving only for the shortest route.

In the literature model, the obtained optimal plan assumes that the shortest routes (i.e., OD1.4, OD2.1, and OD3.2) with bicycle facilities are used by all bicycle flows, whereas the actual flow on these routes is highlighted by the dashed rectangles in Fig. 9. We observe that not all cyclists actually travel via their shortest routes. The shortest routes provided by the literature model, such as OD1.4 and OD3.2, do not even have the greatest route flow. Most cyclists choose other routes with lower route utilities, as shown in Fig. 10.

4.2. Example 2: The Sioux Falls network

We further tested the two proposed solution methods on the Sioux Falls network shown in Fig. 11, which is a widely adopted test network in the transportation network design field, to show the performance of the proposed solution methods and for comparison with the exact method based on enumeration. The bold lines in Fig. 11 represent the candidate links for bike path construction.

The link lengths, unit length construction costs, and OD demands in the example are listed in Tables 10 and 11. The dashes in Table 10 indicate that the corresponding link does not appear in $\widetilde{A}$. The investment budget is set to 4 units. The initial route set of each OD pair was formed by the k-shortest routes, which was generated with the link-elimination method (Bekhor et al., 2001; Long et al., 2010). This method successively removes the middle link on the shortest route to find the shortest route in the remaining network. k is first set to 3, and we then test the sensitivity of the parameter k in this example.

Table 12 lists the initial utility of each route. The value of φ is the same as that in Example 1. To obtain solutions via enumeration within a reasonable time, a randomly selected set of 19 candidate links was used for bike path construction, which is $\widetilde{A}$ = {1, 7, 16, 28, 31, 32, 33, 37, 38, 39, 46, 47, 49, 50, 53, 54, 57, 59, 71}, and thus the total number of solutions is 2^{19} = 524288. For the matheuristic, initial p_{slct} = 0.8, χ = 5, κ = 3, μ_1 = 0.8, μ_2 = 0.9, μ_3 = 1, $C_{fail}^{\max}$ = 5, and $C_{succ}^{\max}$ = 3. The method of generating $I_0 = \{\mathbf{y}_1, \mathbf{y}_2, ...\mathbf{y}_{n_0}\}$ follows remark 1 in Section 3.2.1. Here, we obtained an initial feasible $\mathbf{y}_1$ = (0,1,0,0,0,0,0,0,1,0,0,0,1,1,0,1,0,1,0,1).

Table 13 lists the results from the two proposed solution methods and those from the enumeration method. The second column lists the number of breakpoints (including endpoints) adopted for the feasible region of each U_p^w and α^w. For a fair comparison, the same number of breakpoints was used in the two proposed methods. The third column shows the objective function value Z_a obtained from each method. The fourth column shows the corresponding real objective function value Z_r obtained by evaluating the final solution of each method by solving the PSL traffic assignment problem. Even though the real objective function values obtained with both solution methods are identical to the optimal objective function value obtained with the enumeration method, the matheuristic possesses a smaller gap between Z_a and Z_r, which is only 0.0043%, as shown in the fifth column. This extremely small gap indicates a

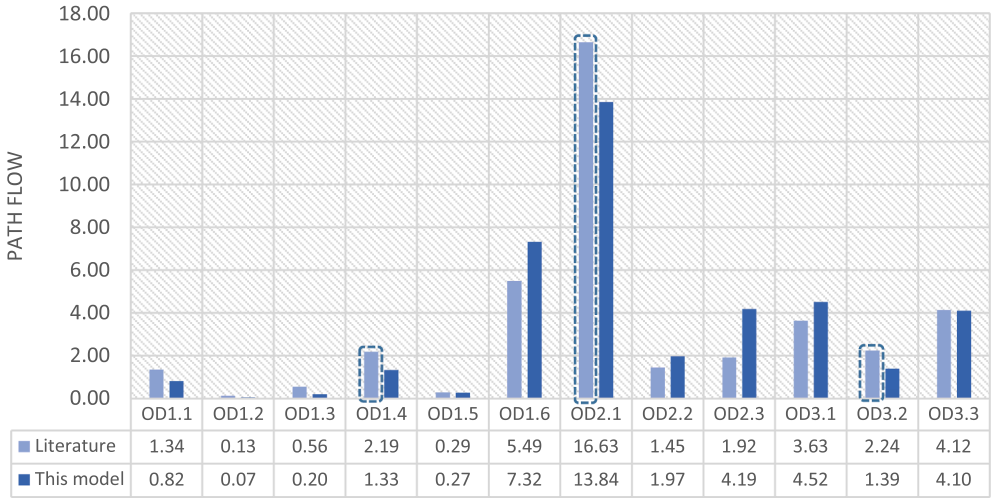
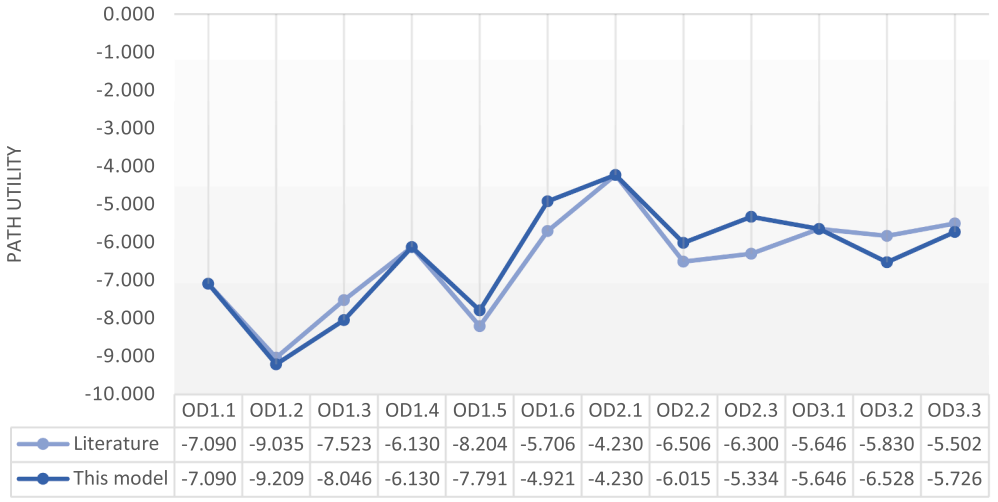


Fig. 9. Comparison of the route flow under the two optimal plans.



	OD1.1	OD1.2	OD1.3	OD1.4	OD1.5	OD1.6	OD2.1	OD2.2	OD2.3	OD3.1	OD3.2	OD3.3
Literature	-7.090	-9.035	-7.523	-6.130	-8.204	-5.706	-4.230	-6.506	-6.300	-5.646	-5.830	-5.502
This model	-7.090	-9.209	-8.046	-6.130	-7.791	-4.921	-4.230	-6.015	-5.334	-5.646	-6.528	-5.726

Fig. 10. Comparison of the route utility under the two optimal plans.

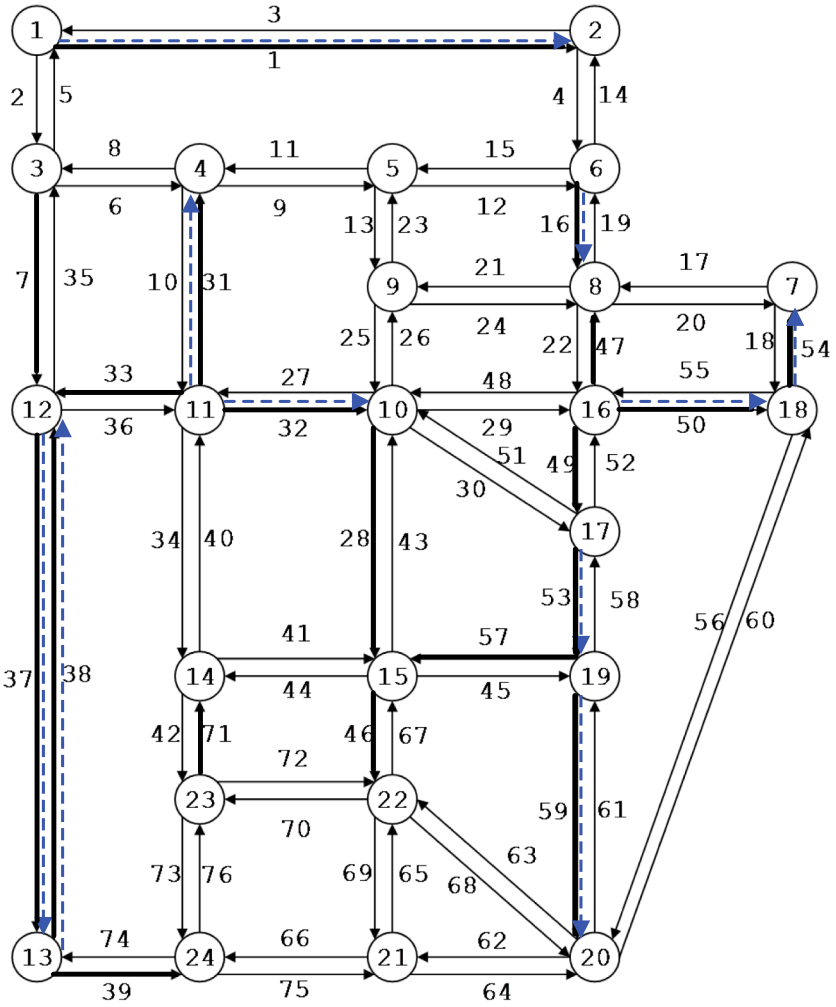


Fig. 11. Sioux Falls network.

Table 10
Link length (miles) and unit length construction cost of the Sioux Falls network.

Link	Length	Cost	Link	Length	Cost	Link	Length	Cost	Link	Length	Cost
1	0.6	0.7	20	0.3	–	39	0.4	1.3	58	0.2	–
2	0.4	–	21	1	–	40	0.4	–	59	0.4	1.5
3	0.6	–	22	0.5	–	41	0.5	–	60	0.4	–
4	0.5	–	23	0.5	–	42	0.4	–	61	0.4	–
5	0.4	–	24	1	–	43	0.6	–	62	0.6	–
6	0.4	–	25	0.3	–	44	0.5	–	63	0.5	–
7	0.4	1.7	26	0.3	–	45	0.4	–	64	0.6	–
8	0.4	–	27	0.5	–	46	0.4	1.6	65	0.2	–
9	0.2	–	28	0.6	1.7	47	0.5	1.5	66	0.3	–
10	0.6	–	29	0.5	–	48	0.5	–	67	0.4	–
11	0.2	–	30	0.8	–	49	0.2	1.4	68	0.5	–
12	0.4	–	31	0.6	0.8	50	0.3	1.6	69	0.2	–
13	0.5	–	32	0.5	0.9	51	0.8	–	70	0.4	–
14	0.5	–	33	0.6	1.1	52	0.2	–	71	0.4	1.7
15	0.4	–	34	0.4	–	53	0.2	0.7	72	0.4	–
16	0.2	0.9	35	0.4	–	54	0.2	1.8	73	0.2	–
17	0.3	–	36	0.6	–	55	0.3	–	74	0.4	–
18	0.2	–	37	0.3	1	56	0.4	–	75	0.3	–
19	0.2	–	38	0.3	1.5	57	0.4	1.2	76	0.2	–

Table 11
OD demand for the Sioux Falls network example.

Index	Origin	Destination	Demand	Index	Origin	Destination	Demand
1	1	7	34	12	12	10	27
2	1	23	3	13	13	4	5
3	2	13	6	14	13	6	12
4	2	17	4	15	16	18	10
5	4	11	9	16	17	5	9
6	6	1	14	17	17	20	13
7	8	15	10	18	17	23	14
8	10	22	13	19	18	6	16
9	11	5	20	20	20	7	11
10	11	13	1	21	22	4	15
11	12	2	6	22	22	18	6

Table 12
Initial utility of each of the k-shortest routes of each OD pair with k = 3.

OD index	Initial utility of each route			OD index	Initial utility of each route		
	1	2	3		1	2	3
1	−3.7536	−4.7124	−6.7524	12	−2.4276	−4.3248	−4.3044
2	−4.1412	−5.0592	−5.0592	13	−2.6316	−3.3660	−4.6920
3	−3.9372	−5.2632	−6.9564	14	−4.1412	−4.8756	−5.0592
4	−3.3864	−5.6304	−6.7524	15	−0.7548	−2.4480	−3.0192
5	−1.3056	−3.1824	−3.5700	16	−3.2028	−3.5700	−3.5496
6	−2.4276	−3.3864	−7.1196	17	−1.5096	−2.2644	−3.9576
7	−3.2028	−3.5496	−4.1412	18	−3.3864	−3.5700	−4.3248
8	−2.2440	−4.1412	−3.9372	19	−1.8972	−2.4480	−4.6920
9	−1.8768	−2.9988	−3.7536	20	−1.5096	−3.4068	−5.6304
10	−2.0604	−3.3864	−3.9372	21	−4.1208	−4.3044	−4.8960
11	−3.1824	−4.5084	−6.3648	22	−2.0604	−2.8152	−3.7740

Table 13
Results of the Sioux Falls network example.

Method	Avg. no. of breakpoints	Z_a	Z_r	Gap	Total time	MILP time
Global optimization	7	417.6156	418.3352	0.1720%	4.40	4.40
Matheuristic	7	418.3174	418.3352	0.0043%	1.18	0.53
Enumeration	–	–	418.3352	–	521.86	–

Table 14
Objective function value and optimal construction plan with various values of k.

k-shortest route	Objective function value when do nothing	Objective value with $B = 4$	Optimal plan with $B = 4$
$k = 1$	493.3578	341.2055	(1,16,31,32,46,49,50,53,54,57)
$k = 3$	558.9039	418.3352	(1,16,31,32,37,38,50,53,54,59)
$k = 6$	584.8034	443.7549	(1,16,31,32,38,49,50,53,54,59)

very fine approximation from the MILP for P3 to the original nonlinear model via this solution method. The total computational time (in seconds) and the time spent on solving the MILP involved (in seconds) are also listed in the last two columns of Table 13. Relative to the global optimization method, the proposed matheuristic required less computational time with the same number of breakpoints, especially less time spent on solving the MILP, which is 11.99% of that of the first solution method. Relative to the enumeration method, both proposed solution methods greatly decrease the computation time by at least 99.16% to obtain the same global optimum.

It should be noted that the number of routes adopted to form the route set is very important. Table 14 illustrates the change in the value of k in the route set on the final objective value and optimal construction plan. We find that as k increases, the objective function values gradually increase when the given budget is 4 units and when no bike paths are introduced on the network (i.e., doing nothing), probably because the cyclists' route choice behavior model takes the form of PSL (a type of stochastic user equilibrium). Hence, in general, the larger the value of k, the larger the objective function value obtained. The optimal plan differs with the use of various k values. However, only three links differ between the optimal plans of $k = 1$ and $k = 3$, and only one link differs between the optimal plans of $k = 3$ and $k = 6$.

4.3. Example 3: The Anaheim network

The Anaheim network, as shown in Fig. 12, is used in this subsection to compare the performance of the two proposed methods in large network applications. It has 416 nodes and 1828 links (the number of links is doubled to transform the undirected links to directed ones). We randomly selected 113 candidate links for the construction of bike paths and randomly set their unit construction costs. We assumed that only 10 major OD demands existed, which were also randomly selected. The network information can be found on the website of Bar-Gera (2019). The value of φ is the same as that in Example 1. For the matheuristic, $\chi = 0.25\ |I_n|$ and other parameters follow those in Example 2 and k is set to 3. The method of generating $I_0 = \{\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_{n_0}\}$ follows remark 1 in Section 3.2.1.

Table 15 lists the construction plans and the corresponding budget, and Table 16 shows the network performance of these plans from the two proposed solution methods and the surrogate-model–based algorithm. The three solution methods provide three construction plans that differ greatly in terms of the number of bike paths to be constructed and the specific links with bike paths. The

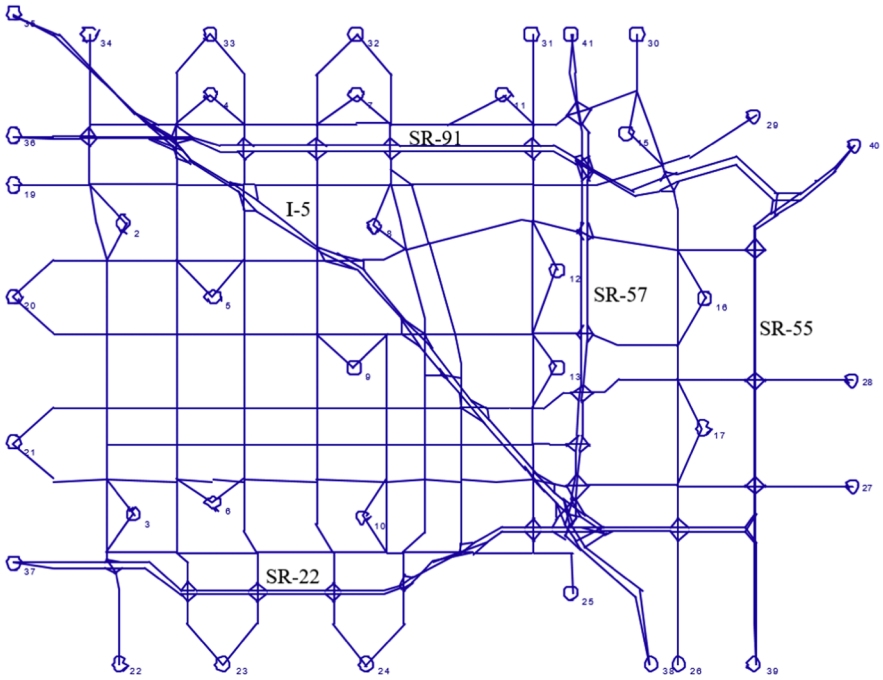


Fig. 12. Anaheim network.

Table 15
Links with bike paths built and budgets used by various methods.

Solution method	Surrogate model–based algorithm	Global optimization	Matheuristic
Link No.	214	214	214
	274	275	275
	275	413	413
	413	544	496
	496	788	544
	544	1166	646
	652	1172	788
	788	1176	1166
	1078	1180	1172
	1115	1194	1176
	1180	1195	1180
	1194	1324	1181
	1195	1443	1194
	1324	1453	1324
	1348	1458	1443
	1354	1506	1453
	1443	1557	1458
	1453	1564	1557
	1461	1567	1564
	1506		1567
	1560		
	1567		
Budget used	9.9987	9.9729	9.9888

Table 16
Results of the Anaheim network example.

Algorithm	Avg. no. of breakpoints	Z_a	Z_r	Gap	Total time	MILP time
Do nothing	/	/	4484.52	/	/	/
Surrogate model based algorithm	/	/	4369.31	/	25.36	/
Global optimization	11	4335.80	4348.90	0.3012%	3698.50	3698.50
Matheuristic	11	4345.19	4345.29	0.0023%	129.67	104.31

global optimization method provides a construction plan with the lowest investment budget. All three methods can provide the final solutions with their real objective function values Z_r that are similar. However, the solution obtained from the matheuristic, which combines the surrogate-model–based algorithm and the global optimization method, is slightly better than that obtained from the global optimization method and that from the surrogate-model–based algorithm. This means that the bike network plan obtained from the matheuristic can provide the best network performance among the three results, which is indeed important in practical application because the plan applied will affect cycling for decades and should thus be as good as possible. Moreover, we note that the calculation time needed for the matheuristic is much shorter than that of global optimization but longer than the surrogate-model–based solution algorithm because it solves the embedded mixed-integer linear program.

As with the result of Example 2, the matheuristic provided a final solution with a smaller gap between Z_a and Z_r. The better solution and smaller gap may have occurred because the surrogate-model–based method removed part of the feasible region that likely does not contain an optimal solution, which makes the MILP a more precise approximation for the original model. In addition to finding a better solution, with the same number of breakpoints, the matheuristic also provided an optimal solution in 129.67 s, which represents a decrease of 96.49% in the computation time relative to that of the global optimization method. The matheuristic saved considerable computation time by using a surrogate model to approximate the original complex problem and to reduce the feasible region of the problem. The time spent on solving the MILP problem also decreased by 97.18% due to the reduction in the feasible region. Hence, the matheuristic is much more efficient in solving the bike network design model, especially with a large network.

Table 17 shows the effects of the value of χ on the solution determined by the matheuristic when the other parameters and

Table 17
Effect of the value of χ on the final solution.

χ	Avg. no. of breakpoints	Z_a	Z_r	Gap	MILP time
$0.20\|I_n\|$	11	4345.8249	4345.9075	0.0019%	30.96
$0.25\|I_n\|$	11	4345.1879	4345.2897	0.0023%	104.31
$0.30\|I_n\|$	11	4345.1533	4345.2897	0.0031%	222.15

settings remain the same. Three values of χ were tested, where $|I_n|$ is the number of evaluated points in the set I_n. In the three tests, with a larger value of χ, the solutions found are generally better while the gap between Z_a and Z_r increases and the MILP computation time also increases due to the larger feasible region. The results are reasonable and consistent with the analysis in the solution methods section. However, a comparison of the results in Table 17 with those in Table 16 shows that even with a smaller χ, the matheuristic can provide a better solution than the global optimization method within a much shorter computation time.

5. Conclusions

In this study, we propose a mixed-integer nonlinear nonconvex model to formulate the bike path network design problem. This model captures cyclists' route choice behavior with the PSL model, in which route utilities can capture many factors that affect the route choices of cyclists. A global optimization method is proposed to solve the reformulated model to obtain the best bike path construction plan with the given investment budget. Considering the application to large network instances, a novel matheuristic is also developed by incorporation of a surrogate-model–based algorithm in the global optimization method.

The numerical experiments show several interesting conclusions. (1) The model and the solution methods can effectively provide optimal bicycle network design plans for an existing urban road network. The enumeration results of both the nine-node network and the Sioux Falls network show that exact solutions are found by both proposed solution methods. (2) The results of the Anaheim network show that the novel matheuristic can provide a better solution while reducing the computation time by 96.49% from the global optimization method. (3) With a very small initial budget, the number of links with bike paths generally increases as the budget increases, but there is an upper bound for construction investment, and exceeding this limit achieves no better performance for the whole network. (4) The OD pairs derive unequal benefits from new bicycle facility construction; some OD pairs derive a much greater increase in utility than others. (5) With consideration of cyclists' realistic route choice, the proposed model shows better system performance than the literature model, and the obtained optimal construction plan requires a lower investment budget.

This study opens up at least four future research directions. First, the proposed model does not consider the effects of congestion because cyclists' route choice behavior was shown to be irrelevant to the bike flow due to a recent study in the United States by Broach et al. (2012). Nevertheless, it may be important to capture the effects of congestion in bicycle network design for other areas around the world, such as Amsterdam and Copenhagen. Moreover, the PSL model may not be suitable for cities in which the cycling level or the traffic laws and rules differ from those in US cities. Therefore, more studies should focus on the development of the cyclists' route choice behavior model from real survey data from other regions. The proposed model and solution method can then be extended to capture the specific situation, such as the effects of congestion, for the cities considered. Second, the proposed solution methods can be applied to solve other network design problems, even those with user equilibrium, as long as the model is reformulated, as we did to reform (P1) and (P2). We plan to apply these methods to solve other network design problems. Third, if the planner plans bike routes that consist only of bike paths so that cyclists can always ride along bike paths from their origins to their destinations, the problem changes to determination of which routes between OD pairs should be selected for construction of bike paths. In this way, the decision changes to whether to construct bike paths along routes instead of constructing bike paths along road links. We plan to address this interesting problem in a future study. Finally, we plan to extend this study to consider other travel modes and the mode split to include interaction between bicycles and other transportation modes.

Acknowledgments

The authors are grateful to the two reviewers for their constructive comments. The research was jointly supported by the National Natural Science Foundation of China (71771194, 71801067, 71431003), the Natural Science Foundation of Anhui Province (1808085MG210) and China Postdoctoral Science Foundation (2018M642514). The authors are grateful to the two reviewers for their constructive comments.

References

Abdulaal, M., LeBlanc, L.J., 1979. Continuous equilibrium network design models. Transp. Res. Part B 13 (1), 19–32.
An, K., Lo, H.K., 2015. Robust transit network design with stochastic demand considering development density. Transp. Res. Part B 81 (Part 3), 737–754.
Aultman-Hall, L., Hall, F., Baetz, B., 1997. Analysis of bicycle commuter routes using geographic information systems: implications for bicycle planning. Transp. Res. Rec.: J. Transp. Res. Board 1578, 102–110.
Bar-Gera, H., 2019. Transportation network test problems. < http://www.bgu.ac.il/~bargera/tntp/ > (access on 4 April 2019).
Bekhor, S., Ben-Akiva, M., Ramming, M.S., 2001. Route choice: Choice set generation and probabilistic choice models. In: Proceedings of the 4th TRISTAN conference, Azores, Portugal.
Bekhor, S., Ben-Akiva, M.E., Ramming, M.S., 2006. Evaluation of choice set generation algorithms for route choice models. Ann. Oper. Res. 144 (1), 235–247.
Bekhor, S., Toledo, T., 2005. Investigating path-based solution algorithms to the stochastic user equilibrium problem. Transp. Res. Part B 39 (3), 279–295.
Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. In: Hall, R.W. (Ed.), Handbook of Transportation Science. Springer, US, Boston, MA, pp. 5–34.
Breckman, J., 1956. Encoding circuit, US Patent 2 733 432.
Broach, J., Dill, J., Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transp. Res. Part A 46 (10), 1730–1740.
Broach, J., Gliebe, J., Dill, J., 2010. Calibrated labeling method for generating bicyclist route choice sets incorporating unbiased attribute variation. Transp. Res. Rec.: J. Transp. Res. Board 2197, 89–97.
Buehler, R., Dill, J., 2016. Bikeway networks: a review of effects on cycling. Transp. Rev. 36 (1), 9–27.
Dial, R.B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. Transp. Res. 5 (2), 83–111.
Dill, J., 2009. Bicycling for transportation and health: The role of infrastructure. J. Public Health Policy 30 (1), S95–S110.

Duthie, J., Unnikrishnan, A., 2014. Optimization framework for bicycle network design. J. Transp. Eng. 140 (7), 04014028.
Ekström, J., Rydergren, C., Sumalee, A., 2013. Solving a mixed integer linear program approximation of the toll design problem using constraint generation within a branch-and-cut algorithm. Transportmetrica A: Transp. Sci. 10 (9), 791–819.
Ekström, J., Sumalee, A., Lo, H.K., 2012. Optimizing toll locations and levels using a mixed integer linear approximation approach. Transp. Res. Part B 46 (7), 834–854.
Farahani, R.Z., Miandoabchi, E., Szeto, W.Y., Rashidi, H., 2013. A review of urban transportation network design problems. Eur. J. Oper. Res. 229 (2), 281–302.
Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. Transp. Res. Part B 41 (3), 363–378.
Gao, Z., Wu, J., Sun, H., 2005. Solution algorithm for the bi-level discrete network design problem. Transp. Res. Part B 39 (6), 479–495.
Goodno, M., McNeil, N., Parks, J., Dock, S., 2013. Evaluation of innovative bicycle facilities in Washington, D.C. Transp. Res. Rec.: J. Transp. Res. Board 2387, 139–148.
Han, S., 2007. A route-based solution algorithm for dynamic user equilibrium assignments. Transp. Res. Part B 41 (10), 1094–1113.
He, X., Chen, X., Xiong, C., Zhu, Z., Zhang, L., 2017. Optimal time-varying pricing for toll roads under multiple objectives: a simulation-based optimization approach. Transp. Sci. 51 (2), 412–426.
Hood, J., Sall, E., Charlton, B., 2011. A GPS-based bicycle route choice model for San Francisco, California. Transp. Lett. 3 (1), 63–75.
Howard, C., Burns, E., 2001. Cycling to work in Phoenix: Route choice, travel behavior, and commuter characteristics. Transp. Res. Rec.: J. Transp. Res. Board 1773, 39–46.
Kepaptsoglou, K., Karlaftis, M., 2009. Transit route network design problem. J. Transp. Eng. 135 (8), 491–505.
Krizek, K.J., El-Geneidy, A., Thompson, K., 2007. A detailed analysis of how an urban trail system affects cyclists' travel. Transportation 34 (5), 611–624.
Löfberg, J., 2004. Yalmip: A toolbox for modeling and optimization in Matlab. In: Proceedings of the 2004 IEEE International Symposium on Computer Aided Control Systems Design, Taipei, Taiwan, pp. 284–289.
LeBlanc, L.J., 1975. An algorithm for the discrete network design problem. Transp. Sci. 9 (3), 183–199.
Li, C., Yang, H., Zhu, D., Meng, Q., 2012. A global optimization method for continuous network design problems. Transp. Res. Part B 46 (9), 1144–1158.
Liu, H., Wang, D.Z.W., 2015. Global optimization method for network design problem with stochastic user equilibrium. Transp. Res. Part B 72, 20–39.
Liu, H., Wang, D.Z.W., 2016. Modeling and solving discrete network design problem with stochastic user equilibrium. J. Adv. Transp. 50 (7), 1295–1313.
Liu, H., Wang, D.Z.W., 2017. Locating multiple types of charging facilities for battery electric vehicles. Transp. Res. Part B 103, 30–55.
Long, J., Gao, Z., Zhang, H., Szeto, W.Y., 2010. A turning restriction design problem in urban road networks. Eur. J. Oper. Res. 206 (3), 569–578.
Lu, X.-S., Liu, T.-L., Huang, H.-J., 2015. Pricing and mode choice based on nested logit model with trip-chain costs. Transp. Policy 44, 76–88.
Luathep, P., Sumalee, A., Lam, W.H.K., Li, Z.-C., Lo, H.K., 2011. Global optimization method for mixed transportation network design problem: a mixed-integer linear programming approach. Transp. Res. Part B 45 (5), 808–827.
Lusk, A.C., Furth, P.G., Morency, P., Miranda-Moreno, L.F., Willett, W.C., Dennerlein, J.T., 2011. Risk of injury for bicycling on cycle tracks versus in the street. Injury Prevent. 17 (2), 131–135.
Mesbah, M., Thompson, R., Moridpour, S., 2012. Bilevel optimization approach to design of network of bike lanes. Transp. Res. Rec.: J. Transp. Res. Board 2284, 21–28.
Miandoabchi, E., Farahani, R.Z., Dullaert, W., Szeto, W.Y., 2012. Hybrid evolutionary metaheuristics for concurrent multi-objective design of urban road and public transit networks. Networks Spatial Econ. 12 (3), 441–480.
Nayan, A., Wang, D.Z.W., 2017. Optimal bus transit route packaging in a privatized contracting regime. Transp. Res. Part A 97, 146–157.
Pattnaik, S., Mohan, S., Tom, V., 1998. Urban bus transit route network design using genetic algorithm. J. Transp. Eng. 124 (4), 368–375.
Pucher, J., Buehler, R., 2006. Why Canadians cycle more than Americans: a comparative analysis of bicycling trends and policies. Transp. Policy 13 (3), 265–279.
Pucher, J., Buehler, R., 2008. Making cycling irresistible: Lessons from the Netherlands, Denmark and Germany. Transp. Rev. 28 (4), 495–528.
Regis, R.G., 2011. Stochastic radial basis function algorithms for large-scale optimization involving expensive black-box objective and constraint functions. Comput. Oper. Res. 38 (5), 837–853.
Regis, R.G., Shoemaker, C.A., 2007. A stochastic radial basis function method for the global optimization of expensive functions. INFORMS J. Comput. 19 (4), 497–509.
Smith, H.L., Haghani, A., 2012. A mathematical optimization model for a bicycle network design considering bicycle level of service. In: Transportation Research Board 91st Annual Meeting, Washington DC.
Snizek, B., Nielsen, T.A.S., Skov-Petersen, H., 2013. Mapping bicyclists' experiences in Copenhagen. J. Transp. Geogr. 30, 227–233.
Szeto, W., Jiang, Y., Wang, D., Sumalee, A., 2015. A sustainable road network design problem with land use transportation interaction over time. Networks Spatial Econ. 15 (3), 791–822.
Szeto, W.Y., Jiang, Y., 2014. Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. Transp. Res. Part B 67, 235–263.
Szeto, W.Y., Lo, H.K., 2006. Transportation network improvement and tolling strategies: the issue of intergeneration equity. Transp. Res. Part A 40 (3), 227–243.
Todd, M.J., 1977. Union jack triangulations. In: Karamardian, S., García, C.B. (Eds.), Fixed points: Algorithms and Applications. Academic Press, New York, pp. 315–336.
Vielma, J., Nemhauser, G., 2011. Modeling disjunctive constraints with a logarithmic number of binary variables and constraints. Math. Program. 128 (1), 49–72.
Vielma, J.P., Ahmed, S., Nemhauser, G., 2010. Mixed-integer models for nonseparable piecewise-linear optimization: unifying framework and extensions. Oper. Res. 58 (2), 303–315.
Wang, D.Z.W., Liu, H., Szeto, W.Y., 2015. A novel discrete network design problem formulation and its global optimization solution algorithm. Transp. Res. Part E 79, 213–230.
Wang, D.Z.W., Lo, H.K., 2010. Global optimum of the linearized network design problem with equilibrium flows. Transp. Res. Part B 44 (4), 482–492.
Wang, G., Gao, Z., Xu, M., Sun, H., 2014. Models and a relaxation algorithm for continuous network design problem with a tradable credit scheme and equity constraints. Comput. Oper. Res. 41, 252–261.
Wang, S., Meng, Q., Yang, H., 2013. Global optimization methods for the discrete network design problem. Transp. Res. Part B 50, 42–60.
Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proc. Inst. Civil Eng. Part II 1, 325–378.
Winters, M., Teschke, K., Grant, M., Setton, E., Brauer, M., 2010. How far out of the way will we travel? Built environment influences on route selection for bicycle and car travel. Transp. Res. Rec.: J. Transp. Res. Board 2190, 1–10.
Wu, J., Liu, M., Sun, H., Li, T., Gao, Z., Wang, D.Z.W., 2015. Equity-based timetable synchronization optimization in urban subway network. Transp. Res. Part C 51 (Supplement C), 1–18.
Yang, H., Bell, M.G.H., 1998. Models and algorithms for road network design: a review and some new developments. Transp. Rev. 18 (3), 257–278.
Yang, H., Huang, H.J., 2005. Mathematical and Economic Theory of Road Pricing. Emerald Group Publishing, UK.
Zhang, X., Huang, H.-J., Zhang, H.M., 2008. Integrated daily commuting patterns and optimal road tolls and parking fees in a linear city. Transp. Res. Part B 42 (1), 38–56.