Transportation Research Part C 125 (2021) 103002

Contents lists available at ScienceDirect

Transportation Research Part C

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

Active learning for multi-objective optimal road congestion pricing considering negative land use effect

Shaopeng Zhong [a], Yunhai Gong [a], Zhijian Zhou [a], Rong Cheng [c], Feng Xiao [b,*]

[a] School of Transportation and Logistics, Dalian University of Technology, Dalian 116024, China
[b] School of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, China
[c] DTU Management, Department of Technology, Management, and Economics, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Keywords:
Road congestion pricing
Negative land use effect
Integrated land use and transportation model
Probabilistic surrogate model
Multi-objective Bayesian optimization
Simulation-based optimization

ABSTRACT

The road congestion pricing policy is implemented to alleviate traffic congestion and improve the efficiency of the transportation system during peak hours. However, the negative land use effect caused by this policy could not be ignored. How to design the optimal congestion toll that can not only ensure its positive effect on the transportation system but also reduce its negative effect on land use is an urgent problem to be solved. Given this, this paper proposed a multi-objective bi-level programming road congestion pricing model based on the integrated land use and transportation model to optimize the regional average accessibility, regional average land use diversity, and regional total flow time. Since the proposed problem is NP-hard, this paper innovatively proposed an active learning optimization algorithm based on multi-objective Bayesian optimization, which improves the computation efficiency of the bi-level programming model by automatically finding the next sampling point (candidate solution) according to the probability information. An empirical analysis of Jiangyin City demonstrated the effectiveness of the proposed approach in coordinating the relationship between land use and transportation and alleviating the negative land use effect caused by road congestion pricing. Moreover, the algorithm proposed in this paper can also be used to solve other transportation-related black box problems with high computation complexity.

1. Introduction

Traffic congestion has been a serious challenge for many cities around the world. It not only increases the travel time and reduces the efficiency of the transportation system, but also exerts a negative impact on the regional economy and environment, which impedes the sustainable development of the region (Sun et al., 2017, 2018; Xiao et al., 2019; Zhong et al., 2020a; Zhang et al., 2021). Economists have pointed out that road congestion pricing is an effective and sustainable economic means to manage urban traffic congestion (Brownstone and Small, 2005). By imposing road congestion pricing on cars traveling during peak hours, the travelers can realize the actual social costs of their travel so that they can use road traffic resources more rationally (Yang and Bell, 1997; Bhat and Castelar, 2002; Yin and Lawphongpanich, 2009; Anas, 2014).

To maximize the effect of the road congestion pricing policy, the government should clarify and make efforts to alleviate its

* Corresponding author.
E-mail addresses: szhong@dlut.edu.cn (S. Zhong), gongyunhai@mail.dlut.edu.cn (Y. Gong), zhouzhijian@mail.dlut.edu.cn (Z. Zhou), roch@dtu.dk (R. Cheng), xiaofeng@swufe.edu.cn (F. Xiao).

https://doi.org/10.1016/j.trc.2021.103002
Received 9 July 2020; Received in revised form 2 January 2021; Accepted 29 January 2021
Available online 23 February 2021
0968-090X/© 2021 Elsevier Ltd. All rights reserved.

negative effect on urban systems. Experiences from the cities that have implemented the road congestion pricing policy indicate that this policy does exert negative effects on regional land use systems while mitigating traffic congestion (Banister, 2002; Arentze and Timmermans, 2007; Tillema et al., 2010; Zhong et al., 2015). The negative land use effect of road congestion pricing will lead to a decrease in regional accessibility and land use diversity (Whitehead, 2002; Quddus et al., 2007; Zhong and Bushell, 2017). Due to the lack of theoretical analysis and optimization methods, the existing road congestion pricing policies usually improve the efficiency of the transportation system at the cost of sacrificing land use performance.

Given this, it is necessary to design an optimal road congestion pricing scheme to guarantee its positive impact on the transportation system while alleviating its negative land use effect. To achieve this goal, two problems should be settled: the first one is to quantify the impact of road congestion pricing on land use; the other is to alleviate the negative land use effects of the road congestion pricing policy. The existing research will be reviewed from the above two aspects.

1.1. Literature review

Three prevailing methodologies have been adopted to investigate the impact of road congestion pricing on land use: empirical analysis (before the introduction of road congestion pricing), ex-post monitoring and analysis (after the introduction of road congestion pricing), and modeling methods. The empirical analysis is conducted before the introduction of road congestion pricing to explore its possible impacts through stated preference surveys or in-depth interviews (Bhat and Castelar, 2002). So far, there are relatively few studies in this area and they are mainly focusing on the impact of road congestion pricing on residential locations, population, business, and jobs. Arentze and Timmermans (2007) did a stated adaption experiment and found that around 11.9% of the respondents would change their working (2.1%) or living (9.8%) location due to road congestion pricing. Tillema et al. (2010) pointed out that people value congestion tolls more than fuel costs when choosing residential locations. Whitehead (2002, 2005) stated that the impact of road congestion pricing on business depends on different causal chains, and the marginal operators and new entrants are most affected by congestion charges after interviewing the British government, real estate developers, business, and academic community.

The ex-post monitoring and analysis method refers to analyzing the effect of road congestion pricing using the actual monitoring data obtained after the implementation of the policy. Road congestion pricing has been implemented in Singapore, London, Oslo, and Stockholm. After a road congestion pricing scheme is introduced, it will take quite a long time to affect the land use system. Thus, only a few empirical studies have been conducted to examine the effects of road congestion pricing on the pattern and diversity of land use and job distribution. Moreover, the observational data are also limited and mainly about the influence of road congestion pricing on business, retail, real estate, and employment. Quddus et al. (2007) investigated the impact of road congestion pricing on retail in central London using a series of econometric models. A comparison of the sales of six retailers inside and outside the toll ring proved that road congestion pricing had a significant negative impact on sales within the toll ring. For example, the sales of the John Lewis Oxford Street store decreased by 8%. It is worth pointing out that the ex-post monitoring and analysis may not be reliable since the results can be influenced by external factors such as the world economic crisis and credit crunch (Transport for London, 2008). Furthermore, since the impact of road congestion pricing on land use will take a long time to be fully reflected, long-term and continuous monitoring is required, which makes it much more difficult to do empirical analysis.

Two types of modeling methods, i.e., independent model and integrated model, are developed to explore the impact of road congestion pricing on land use. The independent model only considers the transportation model and neglects the interactive feedback relationship between land use and transportation, which means that it can only be used to analyze the short-term impact of road congestion pricing on land use. The integrated model or integrated land use and transportation model combines the land use model and the transportation model. Compared with the ex-post monitoring and analysis and the independent model, the integrated land use and transportation model has the following advantages: it can analyze the impact of road congestion pricing for a longer period; it excludes interference from other external uncertainties[1]; it can describe the interactive feedback relationship between land use and transportation. Because of these reasons, this paper adopts the integrated land use and transportation model to analyze the negative land use effect of road congestion pricing. Table 1 summarizes the existing studies on the influence of road congestion pricing (cordon-based road pricing and area-based road pricing) on land use using the integrated model. To sum up, road congestion pricing may decentralize the distribution of jobs, which harms regional land use diversity. Furthermore, it may also cause the boundary effect, which will reduce the accessibility of some areas.

It should be noted that many existing studies have identified the negative land use effects of road congestion pricing based on integrated land use and transportation model. To the best of the authors' knowledge, however, how to optimize the negative land use effects of road congestion pricing has not been effectively resolved. This is because the integrated land use and transportation model is similar to a black box whose computation efficiency is very low, and the analytical solution of the integrated model cannot be obtained. As a result, the existing studies usually combine the integrated model with scenario planning technology, which can only analyze some specific scenarios but cannot obtain the optimal value of a specific objective function, not to mention to consider multiple objectives.

To fill this research gap, this paper proposed a multi-objective bi-level programming model based on the integrated land use and transportation model to determine the optimal road congestion pricing, intending to reduce the negative land use effects on the premise of not (or less) influencing the efficiency of the transportation system (measured by travel time). Based on the equilibrated

[1] It's worth noting that the integrated land use and transportation model adopted in this study, namely the TRANUS model, may not be able to capture broader uncertainties due to the assumption behind its logit-based formulation.

Table 1
Land use effects of road congestion pricing obtained from the integrated models.

Study area and pricing scheme	Authors	Integrated model	Conclusions
Brussels: Cordon-based road pricing in Brussels metropolitan area	Lobe et al. (1998)	TRANUS	In the charging area, with different charging schemes, the change rate of population and that of the number of tertiary industry jobs are −2.0 to +4.0% and −8.0 to +1.0%, respectively.
Edinburgh: Cordon-based road congestion pricing in the city center	Still et al. (1999)	DELTA/START and LUCI	Residential areas are slightly concentrated (0.2%) and jobs are dispersed (−5.5 to −7.0%).
Austin: Area-based road pricing	Gupta et al. (2006)	DRAM-EMPAL/ TransCAD	Jobs are dispersed from the city center to the outsides (up to 2.2%).
Washington, DC: Cordon-based road congestion pricing in the city center	Safirova et al. (2006)	LUSTRE/ START/RELU	The number of jobs within the toll area has decreased. The wages of workers and the welfare benefits have increased.
Jiangyin: Cordon-based road congestion pricing in the city center	Zhong et al. (2015)	TRANUS	Under different toll rates, the number of retail jobs and the population within the toll ring has decreased by 1.32–1.25% and 3.36–3.06%, respectively.
Jiangyin: Cordon-based road congestion pricing in the city center	Zhong and Bushell (2017)	TRANUS	The implementation of road congestion charging will cause the boundary effect, which will greatly reduce the job accessibility of the periphery of the toll ring.

interactive feedback relationship between the land use and transportation systems obtained from the lower-level problem, the upper-level problem is to determine the optimal congestion toll considering multiple objectives. The lower-level problem is the integrated land use and transportation model which describes the equilibrated interactive feedback relationship between the land use and transportation systems under a specific congestion charging scheme given by the upper-level problem.

The difficulties in solving the proposed problem lie in the following two points. First, the proposed problem takes multiple objectives into account since the road congestion pricing policy will exert influences on various aspects, such as population, jobs, land use diversity, accessibility, and travel time. As for the multi-objective problem, there usually does not exist a solution that can optimize every single objective. Therefore, the Pareto optimal solution is adopted by the existing researches (Yang and Zhang, 2002; Zhang et al., 2008; Sumalee et al., 2009; Xiao and Zhang, 2014; Liu et al., 2014). Second, and more importantly, the bi-level programming problem is NP-hard even if both the upper- and lower-level models are linear. The lower-level problem of the proposed model is non-linear, which makes the model more complicated and difficult to be solved. A series of approaches, such as transformation methods (Allende and Still, 2013), fuzzy methods (Pramanik and Roy, 2007), global techniques (Wang et al., 2010), enumeration methods (Thoai et al., 2005), and meta-heuristic algorithms (Sun et al., 2008; Talbi, 2013; Szeto and Jiang, 2014), have been used to solve the bi-level programming problems. In terms of the problem proposed in this paper, since the lower-level integrated model will be frequently called, the computation efficiency of the traditional solution methods will be significantly affected. In addition, because the analytical solution of the lower-level integrated model cannot be obtained, the relationship between the upper-level objective function and decision variables is similar to a “black box”, which brings difficulties to solve the proposed problem.

Some scholars have proposed simulation-based methods to solve this kind of black box problem, which require a huge amount of computational demands and can not be solved analytically (Osorio and Bierlaire, 2013; Osorio and Chong, 2015; Chen et al., 2019). For example, Osorio and Bierlaire (2013) and Osorio and Chong (2015) proposed a metamodel as the surrogate model of the objective function. The model integrates the information of an analytical model and a simulation-based model to improve computational efficiency. In contrast, this study adopts a probabilistic surrogate model, namely Gaussian process model, to establish the relationship between the decision variable (i.e., the road pricing scheme) and the objective function (i.e., various land use and transportation indicators). Compared with the metamodel (an analytical model plus a simulation-based model), the probabilistic surrogate model used in this study is easier to be constructed. Recently, Chen et al. (2019) proposed a simulation-based optimization algorithm to solve the heteroscedasticity stochastic problem. The algorithm considers the uncertainty of metamodel parameters in a Bayesian framework. The distribution of the objective function is estimated by the Bayesian stochastic Kriging model and optimized by a genetic algorithm. It’s worth noting that, since a genetic algorithm is used to optimize the objective function, it still needs to call a large number of lower-level black box problems, which affects the computational efficiency of the model.

To tackle this problem, this paper developed an active learning optimization algorithm based on multi-objective Bayesian optimization, which can reduce the times of calling the lower-level integrated model and determine the upper-level optimal congestion pricing scheme quickly. In contrast with the simulation-based methods with a metamodel model, the surrogate model used in this study is easier to be constructed. Compared with the heuristic algorithm, such as a genetic algorithm, the proposed method is more efficient (in terms of the required number of objective function evaluations) and the quality of the solution is better (Olofsson et al., 2019). This is because, with the help of the probabilistic surrogate model (the Gaussian process), we can treat the black box problem like an analytical function, whilst providing us with a measure of confidence bounds for selecting unexplored optima, which greatly reduces the number of calls to the lower-level black box problem.

Bayesian optimization is a type of gradient-free global optimization algorithm (Shahriari et al., 2015; Rodrigues and Pereira, 2018). Since the objective function is an unknown “black box”, Bayesian optimization usually assumes that the objective function is sampled from a probabilistic surrogate model. By establishing a Gaussian process model based on the training data, Bayesian optimization can capture the features of unknown functions. On this basis, Bayesian optimization searches the potential optimal solution in the unexplored feasible region and determines the next sampling point using an acquisition function. The global optimal value of the objective function is obtained by repeating the above steps.

Researchers have confirmed the effectiveness and practicability of multi-objective Bayesian optimization in solving multi-objective problems and bi-level problems (Shahriari et al., 2015; Feliot et al., 2017; Kieffer et al., 2017). The method has been applied in power amplifier design (Chen et al., 2015), data mining (Golovin et al., 2017), chemical reactor design (Park et al., 2018), and engineering design (Nahvi et al., 2019). However, the application of multi-objective Bayesian optimization in the field of transportation science is rare.

1.2. Objectives and contributions

The objectives of this paper are: (1) to develop a multi-objective bi-level programming model based on the integrated land use and transportation model to quantitatively analyze the negative land used effects of road congestion pricing; (2) to devise an efficient algorithm that can solve the proposed problem; and (3) to coordinate the impacts of road congestion pricing on land use and transportation systems and mitigate the negative effects of this policy by optimizing the congestion tolls according to the principle of Pareto optimality.

The contributions of this paper are summarized as follows.

(1) A multi-objective bi-level programming road congestion pricing model considering the negative land use effects is established.
(2) A series of indicators are proposed to measure the land use effect of road congestion pricing. According to the Pareto optimization principle, by optimizing the toll rate, the negative impact of the road congestion pricing policy on regional land use can be minimized without affecting (or less affecting) the operation efficiency of the transportation system.
(3) An active learning algorithm based on multi-objective Bayesian optimization is designed to solve the proposed model. To the best of the authors' knowledge, this is the first attempt to solve the multi-objective bi-level programming road congestion pricing problem using multi-objective Bayesian optimization.
(4) A case study in Jiangyin City proves that the proposed approach can make the indicators of the land use and transportation systems reach Pareto optimality and alleviate the negative land use effects of road congestion pricing.

The rest of the paper is organized as follows: a bi-level programming model is presented in Section 2. The multi-objective Bayesian optimization algorithm based on Gaussian process is introduced in Section 3. Section 4 verifies the effectiveness of the proposed model and algorithm through an empirical case study. Section 5 concludes the paper.

2. Model formulation

2.1. Problem description

The existing road congestion pricing policy usually improves the transportation system performance at the cost of sacrificing the land use performance. The effects of road congestion pricing on the land use performance and the transportation system performance are neither positively correlated nor completely opposed. Given this, from the perspective of land use and transportation integration, this study determines the optimal road congestion toll by exploring the long-term land use effect of road congestion pricing policy, which can make the regional average accessibility, regional average land use diversity, and regional total flow time reach Pareto optimality. In this paper, Pareto optimality is achieved when a road congestion charging scheme cannot make the regional total flow time better off without making the regional average accessibility or regional average land use diversity worse off.

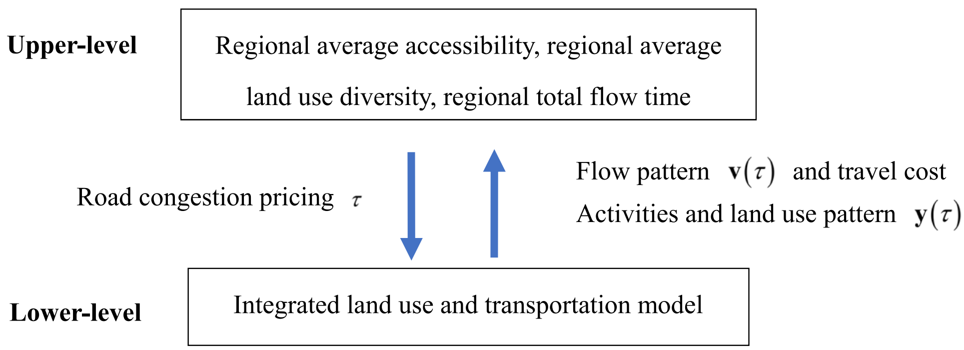

Fig. 1. The framework of the multi-objective bi-level programming road congestion pricing model considering the negative land use effect.

2.2. Model formulation

The study area is divided into a series of traffic analysis zones (TAZs), notated as $Z = \{z^1, z^2, ..., z^M\}$, where M is the number of TAZs. The objective function is notated as $C = \{c^1, c^2, ..., c^W\}$, where W is the number of objective functions. The framework of the multi-objective bi-level programming road congestion pricing model considering the negative land use effect is illustrated in Fig. 1.

The upper-level problem determines the optimal congestion toll that optimizes various objectives (i.e., regional average accessibility, regional average land use diversity, and regional total flow time) based on the equilibrated land use and traffic flow patterns obtained from the lower-level problem. The lower-level problem discovers the interactive feedback relationship between the land use and transportation systems under a given congestion charge scheme and returns the equilibrated traffic flow pattern, travel cost, and land use pattern (i.e., population, jobs, spatial distribution, and the number of different land use types). The upper-level problem can evaluate the conditions of the transportation system as well as the land use performance since the lower-level problem considers the interactive feedback relationship between land use and transportation systems.

2.2.1. The upper-level problem

Two types of objectives are considered in the upper-level problem.

(1) Land use performance indicators – regional average accessibility and regional average land use diversity.
  (a) Regional average accessibility. It refers to the ease of residents reaching a certain activity. Regional accessibility, calculated by Formula (1), depends on the number of certain activities in a region and the travel time between different regions.

$$A_i = \sum_{j \in Z} E_j \cdot I(t_{ij}) = \sum_{j \in Z} E_j [\omega \cdot exp(\gamma + \lambda \cdot t_{ij})], \forall i \in Z \tag{1}$$

Regional average accessibility is calculated by Formula (2):

$$\overline{A} = \sum_{i \in Z} A_i / M \tag{2}$$

where A_i is the accessibility of region i; E_j is the number of an activity in region j; $I(t_{ij})$ is the impedance function of travel time, ω, γ, λ are the parameters of the impedance function; t_{ij} is the minimum travel time between region i and region j; $\overline{A}$ is the regional average accessibility; M is the number of TAZs.

  (b) Regional average land use diversity. This indicator reflects the diversity of land uses within a region. When a variety of land uses exist in a given area, it is expected that many trips originating from that area may have trip destinations within the same area. This study uses Shannon entropy to describe the land use diversity in a region, which is associated with the land use types and quantities of the region. Land use diversity is calculated as follows:

$$D_i = -\frac{\sum_{s=1}^{S_i} p_i^s \cdot lnp_i^s}{lnS_i}, \forall i \in Z \tag{3}$$

Regional average land use diversity is calculated using Formula (4):

$$\overline{D} = \sum_{i \in Z} D_i / M \tag{4}$$

where D_i is the land use diversity of region i; p_i^s is the proportion of the area of land use type s in region i to the total area of land use in this region; S_i is the number of land use types in region i; $\overline{D}$ is the regional average land use diversity.

(2) Transportation performance indicator – regional total flow time

Regional total flow time is calculated by Formula (5):

$$T = \sum_{i \in Z} \sum_{j \in Z, j \neq i} q_{ij} \cdot t_{ij} \tag{5}$$

where T is the regional total flow time; q_{ij} is the traffic flow between region i and region j.

2.2.2. The lower-level problem

The lower-level problem is the integrated land use and transportation model, which is used to determine the equilibrated state between the land use and transportation systems under a given congestion charge. In this study, the TRANUS model, an integrated land use and transportation model based on the logit model, is adopted in the lower-level problem. The TRANUS model determines the complex interactive feedback relationship between the land use and transportation systems, i.e., the spatial location choices, trip generation, trip distribution, modal split, and traffic assignment, according to the principle of maximizing the expected utility. The theoretical basis of the TRANUS model includes random utility theory, gravity and entropy models, spatial microeconomics, and input–output analysis. For more details about the TRANUS model, the readers can refer to de la Barra et al. (1984) and Zhong et al. (2015).

The reasons for choosing the TRANUS model as the lower-level model are that: First of all, TRANUS is a classic integrated land use and transportation model, which has been proved to be able to simulate the complex relationship between urban land use and transportation caused by various land use and transportation policies, such as road congestion pricing (de la Barra et al., 1984; Bandeira et al., 2011; Zhong et al., 2015). Secondly, TRANUS is suitable for the needs of this study. On the one hand, it is difficult to get the analytical solution of the problem based on the TRANUS model, which makes the proposed problem similar to a black box; on the other hand, the TRANUS model requires a lot of computation, which leads to the low efficiency of solving the optimal solution of the problem with the traditional heuristic algorithm. It should be noted that this study intends to propose a general simulation-based active learning optimization framework that does not depend on the properties of the lower-level model. Therefore, other integrated land use and transportation models can also be used as the lower-level model as long as it can reflect the interactive feedback relationship between the land use and transportation systems. In Section 3, we will provide a detailed description of how the lower-level integrated land use and transportation model can analyze the changes in land use and transportation systems caused by road congestion pricing (see Fig. 2).

2.2.3. *Multi-objective bi-level programming model*

Let τ, $\mathbf{v}(\tau)$, $\mathbf{y}(\tau)$, $L(\tau)$, f^w be the congestion toll, the equilibrium traffic flow, the equilibrium land use pattern, the integrated land use and transportation model, and the w^{th} objective function, respectively. The multi-objective bi-level programming model of road congestion pricing problem considering the negative land use effect is presented as model P_0:

$$\max_{\tau}\mathbf{f}(\tau,\mathbf{v}(\tau),\mathbf{y}(\tau)) \tag{6}$$

s.t.

$$\tau\in\Omega \tag{7}$$

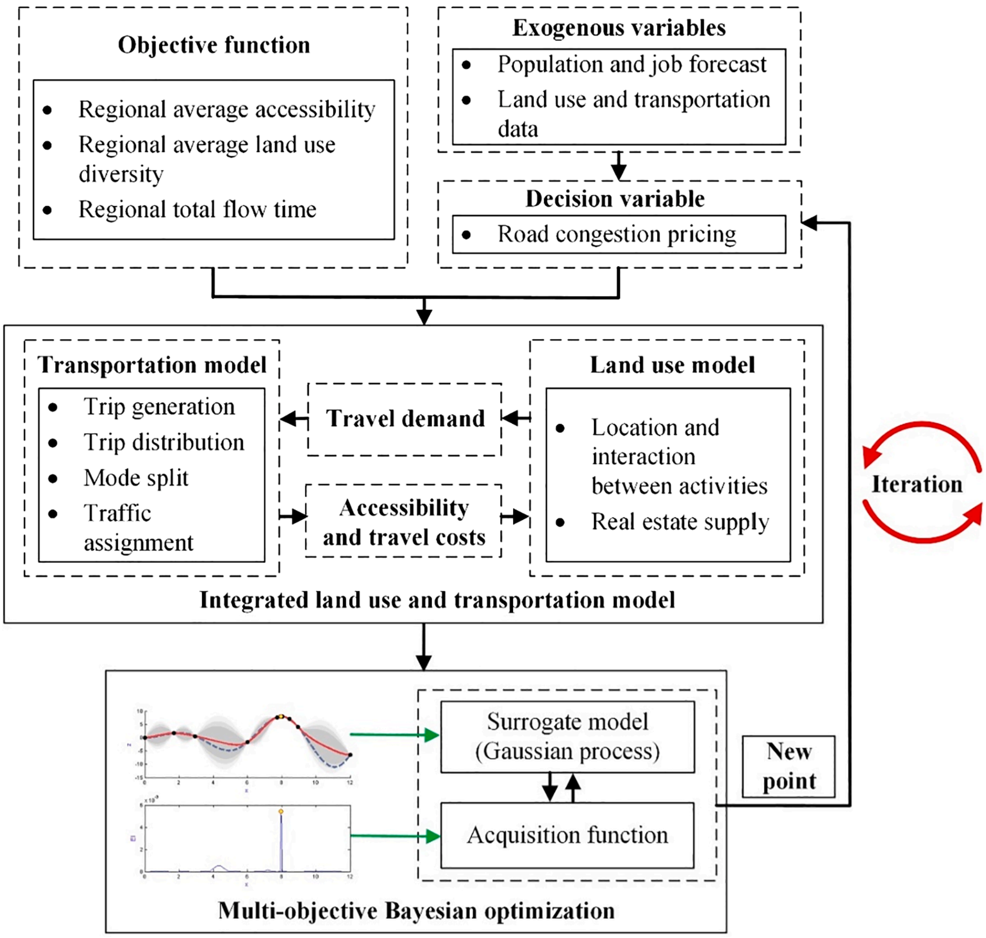

Fig. 2. The optimization procedure of the multi-objective bi-level programming road congestion pricing problem.

$$[\mathbf{v}(\tau), \mathbf{y}(\tau)] = L(\tau) \tag{8}$$

Formula (6) represents a set of objective functions, given by the following formula:

$$\mathbf{f}(\tau, \mathbf{v}(\tau), \mathbf{y}(\tau)) = \{f^w(\tau, \mathbf{v}(\tau), \mathbf{y}(\tau)), \forall w \in C\} \tag{9}$$

Formula (7) limits the solution space of the congestion charge schemes. Formula (8) presents the equilibrium state between the land use and transportation systems obtained by the lower-level integrated land use and transportation model, where $[\mathbf{v}(\tau), \mathbf{y}(\tau)]$ is the equilibrium solution under a given congestion pricing scheme. For ease of expression, in the next part of this paper, $f^w(\tau, \mathbf{v}(\tau), \mathbf{y}(\tau))$ and $\mathbf{f}(\tau, \mathbf{v}(\tau), \mathbf{y}(\tau))$ are denoted as $f^w(\tau)$ and $\mathbf{f}(\tau)$, respectively.

3. An active learning algorithm for the multi-objective bi-level programming model

This study aims to determine the optimal congestion pricing so that the regional average accessibility, regional average land use diversity, and regional total flow time can reach Pareto optimality. Since the proposed problem is NP-hard, an active learning optimization algorithm based on multi-objective Bayesian optimization is devised to reduce the times of calling the lower-level integrated land use and transportation model and improve computation efficiency.

Fig. 2 illustrates the optimization procedure of the multi-objective bi-level programming road congestion pricing problem. The change of the decision variable (congestion charge) will influence the performance of the land use and transportation systems in the region. Specifically, on one hand, the implementation of road congestion pricing will change the short-term travel behavior of travelers (such as travel frequency, travel mode, travel path, etc.) and the operating conditions of the transportation system, and gradually influence the land use system through accessibility (Zhong et al., 2020b). On the other hand, in the long-term, as the impact of road congestion pricing on the land use system increases, road congestion charging policy will also affect the long-term spatial location decisions of both residents and firms, and once again impact travel distribution and travel behavior of residents (Arentze and Timmermans, 2007; Zhong et al., 2015). According to the above process, the land use and transportation systems influence each other until they reach an equilibrium state (Banister, 2002; Zhong et al., 2021). Through the lower-level integrated land use and transportation model, the interaction between land use and transportation systems under the influence of road congestion pricing can be analyzed, and the equilibrium solution and the corresponding objective values can be obtained. According to the value of each objective function, an active learning algorithm based on multi-objective Bayesian optimization is devised to determine the next candidate solution (congestion charge). The above steps are repeated until the Pareto optimal solution is obtained. Bayesian optimization, multivariate Gaussian process, and acquisition function are introduced in the following subsections, based on which the multi-objective Bayesian optimization-integrated land use and transportation model optimizer (MBO-ILUTM optimizer) is proposed.

3.1. Bayesian optimization

Bayesian optimization is a global optimization method for the black box problem, which establishes a surrogate model of the objective function using the prior data and does not rely on the gradient of the objective function. According to the two core parts of Bayesian optimization (i.e., the surrogate model (Gaussian process regression model) and the acquisition function), the next sampling point with the best improvement can be calculated, which allows Bayesian optimization to bypass the black box with high computational complexity (in this paper, it refers to the integrated land use and transportation model) when selecting sampling points. As a result, the computation of Bayesian optimization is less than that of meta-heuristic algorithms (Shahriari et al., 2015). For this reason, this study develops an MBO-ILUTM optimizer to solve the multi-objective bi-level programming road congestion pricing problem.

3.2. Multivariate Gaussian process

The multi-objective bi-level programming road congestion pricing problem is NP-hard so that it is difficult to find the analytical solution to the model. Therefore, this paper uses a surrogate model to fit the relationship between the decision variable and the value of the objective function. The zero noise, zero mean Gaussian process regression model is used as a surrogate model, which is simple and easy to calculate (Rasmussen, 2003).

Taking the objective function w as an example, assume that there are n sampling points $\tau_1, \cdots, \tau_n$ whose objective function values are $f^w(\tau_1), \cdots, f^w(\tau_n)$, respectively, then the multivariate Gaussian distribution is represented as

$$\mathrm{f}^w_{1:n} \sim N(0, \mathbf{K}) \tag{10}$$

where $\mathrm{f}^w_{1:n} = (f^w(\tau_1) \cdots f^w(\tau_n))^{\mathrm{T}}, \forall w \in C; \mathbf{K}_{ij} = k(\tau_i, \tau_j), k(\tau_i, \tau_j)$ is the covariance kernel function. In this study, the common Matérn 5/2 kernel is used as the covariance kernel function:

$$k_{\text{Matérn}\frac{5}{2}}(\tau_i, \tau_j) = \left(1 + \frac{\sqrt{5}\,r}{l} + \frac{5r^2}{3l^2}\right)\exp\left(-\frac{\sqrt{5}\,r}{l}\right) \tag{11}$$

where l is the hyperparameter that determines the curve of the Gaussian process, $r = |\tau_i - \tau_j|$.

The value of l is determined by optimizing the logarithmic form of the marginal likelihood function of the Gaussian process regression model using a gradient method when training the surrogate model (Rasmussen and Williams, 2006).

$$logp\left(\mathrm{f}^{w}_{1:n}|\tau_{1:n}\right) = -\frac{1}{2}\mathrm{f}^{wT}_{1:n}\mathbf{K}^{-1}\mathrm{f}^{w}_{1:n} - \frac{1}{2}log|\mathbf{K}| - \frac{n}{2}log2\pi \tag{12}$$

where $p\left(\mathrm{f}^{w}_{1:n}|\tau_{1:n}\right)$ is the marginal likelihood function of the Gaussian process based on the training set.

Given a sampling point, the distribution of its objective function value predicted by the trained Gaussian process regression model is expressed as

$$f^{w}(\tau_{n+1})\tilde{}N(\mu^{w}(\tau_{n+1}),\sigma^{w}(\tau_{n+1})) \tag{13}$$

The mean and variance are given by Formulas (14) and (15), respectively:

$$\mu^{w}(\tau_{n+1}) = \mathbf{K}^{\mathrm{T}}_{n+1}\mathbf{K}^{-1}\mathrm{f}^{w}_{1:n} \tag{14}$$

$$\sigma^{w}(\tau_{n+1}) = k(\tau_{n+1},\tau_{n+1}) - \mathbf{K}^{\mathrm{T}}_{n+1}\mathbf{K}^{-1}\mathbf{K}_{n+1} \tag{15}$$

where $\mathbf{K}_{n+1} = (k(\tau_{n+1},\tau_1),\dots,k(\tau_{n+1},\tau_n))^{\mathrm{T}}$.

3.3. Acquisition function

The next sampling point is determined by the mean and variance of the acquisition function. The optimal value of the objective function can be obtained with as few iterations as possible by maximizing the acquisition function, as expressed in Formula (16).

$$\tau_{\text{next}} = \underset{\tau\in\Omega}{argmax}\mathrm{AC}(\tau) \tag{16}$$

where $\mathrm{AC}(\cdot)$ is the acquisition function.

The acquisition functions in multi-objective Bayesian optimization are mainly divided into two categories: the scalar-based method and the Pareto optimal-based method. This study adopts the Pareto optimal-based method, which aims to find the Pareto front in the objective function space where there are pairs of non-dominated points and dominated points. The non-dominated points are superior to or equal to the dominated points in either direction. The objective function space is divided into non-dominated area and dominated area by the Pareto front.

Hypervolume-based probability of improvement (HVPoI) (Couckuyt et al., 2014), expected maxmin improvement function (Svenson and Santner, 2016), and Euclidean distance-based expected improvement function (Keane, 2006) are three prevailing Pareto optimal-based methods. After repeated experiments, this study finally decided to use HVPoI as the acquisition function. HVPoI does not require data normalization and can significantly reduce the overall calculation complexity. The detailed calculation steps of this method are given as follows.

(1) Calculate the probability of improvement.

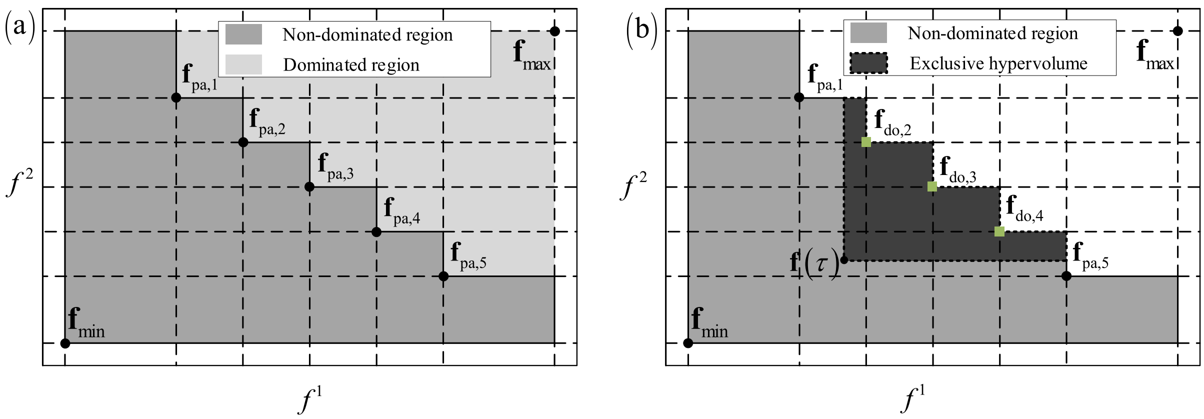

Fig. 3. Illustration of the Pareto front of two objective functions. Note: $\mathbf{f}^{i}_{\text{pa}}$ and $\mathbf{f}^{i}_{\text{do}}$ are the points on the Pareto front and the dominated points, where $i = 1,\dots5$; $\mathbf{f}_{min}$ and $\mathbf{f}_{max}$ denote the lower and upper bound of the hypervolume region. The dark and light shaded areas in figure (a) represent the non-dominated area and the dominated area, respectively. The non-dominated area is the integration area Q based on multi-dimensional space $P_{\mathrm{M}}(\tau)$, and the hypervolume of the dominated area bounded by $\mathbf{f}_{min}$ and $\mathbf{f}_{max}$ is the improvement of the Pareto set $\mathscr{P}$. The entire shaded area is decomposed into several independent rectangular areas by a binary partitioning procedure based on the existing points on the Pareto front. Figure (b) illustrates that the exclusive hypervolume of the objective function value point $\mathbf{f}(\tau)$ relative to the Pareto set $\mathscr{P}$ (the black shaded region) can be computed from the existing cells.

$$P_{\mathrm{M}}(\tau) = \int_{\mathbf{f}(\tau)\in Q} \prod_{w=1}^{W} \phi^{w}[f^{w}(\tau)]df^{w}(\tau) \tag{17}$$

where $P_{\mathrm{M}}(\tau)$ is the probability of a new sampling point located in the non-dominated space. It can also be regarded as the probability that a new input point yields improvement over the Pareto set $\mathscr{P}$. W is the number of the objective functions; Q is the non-dominated part of the objective function space; $\phi^{w}[f^{w}(\tau)]$ is the probability density function of $f^{w}(\tau)$; $\mathbf{f} = \left(f^{1}\cdots f^{W}\right)$.

To simplify the calculation, this study uses the method proposed in Couckuyt et al. (2014) where the integration area Q is decomposed into b (hyper-)rectangular cells, as shown in Fig. 3(a). $P_{\mathrm{M}}(\tau)$ is calculated by the following formula

$$P_{\mathrm{M}}(\tau) = \sum_{a=1}^{b} \pm \prod_{w=1}^{W} \left(\Phi^{w}\left[u_{a}^{w}\right] - \Phi^{w}\left[l_{a}^{w}\right]\right) \tag{18}$$

where u_a^w and l_a^w are the upper bound and lower bound of the w^{th} objective function in region a, respectively; Φ^{w} is the cumulative distribution function of $f^{w}(\tau)$.

(2) Calculate the hypervolume improvement of a Pareto set $\mathscr{P}$ brought by the objective function value point $\mathbf{f}(\tau)$, see Fig. 3(b).

$$H_{\mathrm{imp}}(\mathbf{f}(\tau), \mathscr{P}) = H(\mathscr{P} \cup \{\mathbf{f}(\tau)\}) - H(\mathscr{P}) \tag{19}$$

where $H(\cdot)$ is the hypervolume indicator; H_{imp} is the hypervolume improvement of the objective function value point $\mathbf{f}(\tau)$ to the Pareto set $\mathscr{P}$ and can be defined by a non-negative scalar improvement function, namely,

$$I(\mathbf{f}(\tau), \mathscr{P}) = \begin{cases} H_{imp}(\mathbf{f}(\tau), \mathscr{P}), & \text{if} \quad \mathbf{f}(\tau) \text{ is not dominated by } \mathscr{P} \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

It is worth noting that, in this paper, the hypervolume improvement is not only a standard for selecting sampling points but also an evaluation index for judging the convergence of the algorithm and evaluating the quality of the solution. We will make a detailed explanation with the actual example in Section 4.2.1.

(3) Create HVPoI-based acquisition function.

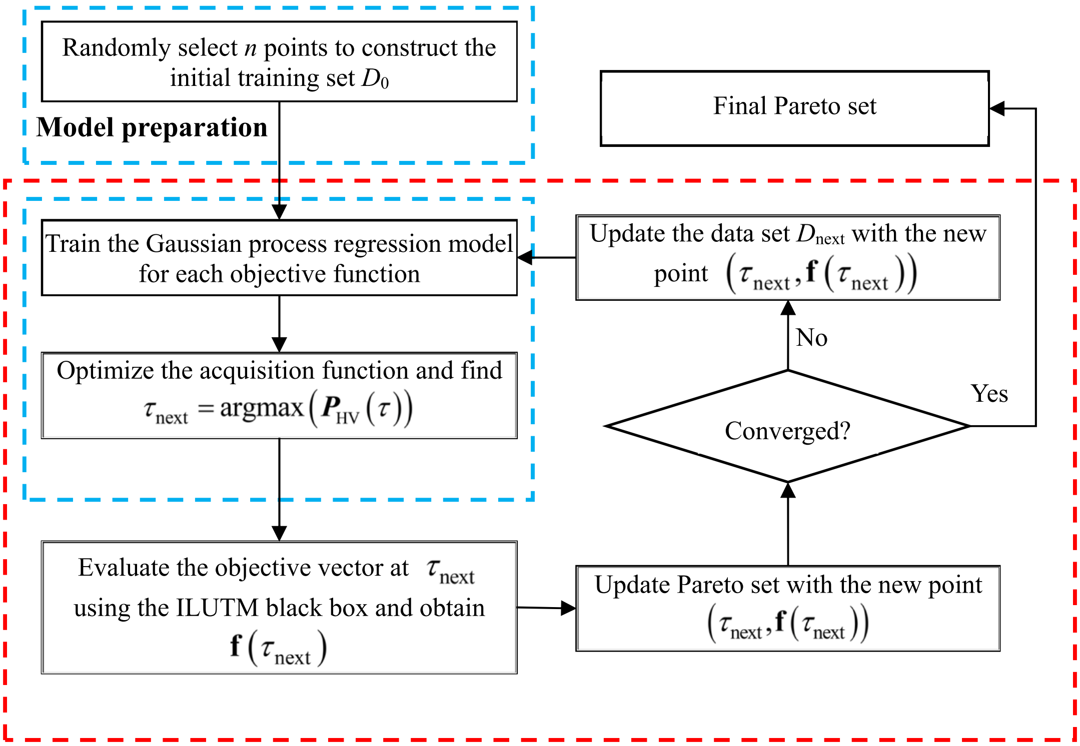

Fig. 4. The procedure of the MBO-ILUTM optimizer.

The HVPoI-based acquisition function $\boldsymbol{P}_{\mathrm{HV}}(\tau)$ can be represented by the product of the improvement function $\boldsymbol{I}(\boldsymbol{\mu}(\tau), \mathscr{P})$ and the improvement probability $\boldsymbol{P}_{\mathrm{M}}(\tau)$.

$$\boldsymbol{P}_{\mathrm{HV}}(\tau) = \boldsymbol{I}(\boldsymbol{\mu}(\tau), \mathscr{P}) \cdot \boldsymbol{P}_{\mathrm{M}}(\tau) \tag{21}$$

where $\boldsymbol{\mu} = \left(\mu^1(\tau), ..., \mu^W(\tau)\right)$.

Similarly, Formula (22) is obtained by dividing the integration area Q into b (hyper-) rectangular cells.

$$\boldsymbol{P}_{\mathrm{HV}}(\tau) = \left(\sum_{a=1}^{b} \pm \mathscr{V}(\boldsymbol{\mu}(\tau), \mathbf{l}_{\mathrm{a}}, \mathbf{u}_{\mathrm{a}}) \right) \cdot \boldsymbol{P}_{\mathrm{M}}(\tau) \tag{22}$$

where $\mathscr{V}(\boldsymbol{\mu}(\tau), \mathbf{l}_{\mathrm{a}}, \mathbf{u}_{\mathrm{a}}) = \begin{cases} \prod_{w=1}^{W} \left(u^w - max\left(l_a^w, \mu^w(\tau)\right)\right), & \text{if} \quad u^w > \mu^w(\tau), \forall w \in C \\ 0, & \text{otherwise} \end{cases}$;$[\mathbf{l}_a, \mathbf{u}_a], \forall a = 1, ..., b$ represents the lower and upper bound of rectangular region a.

3.4. MBO-ILUTM optimizer

An MBO-ILUTM optimizer is developed to solve the multi-objective bi-level programming road congestion pricing problem. The optimization procedure of the MBO-ILUTM optimizer is presented in Fig. 4.

The initial sampling points set, namely, the initial training set is obtained in the model preparation stage. In the Bayesian optimization stage, the training set is used to train the Gaussian process regression model and obtain the mean and variance of the sampling points, based on which the acquisition function is established; then, the next sampling point τ_{next} with the biggest improvement is obtained by optimizing the acquisition function. After determining the sampling point, the ILUTM calculates the objective function value $\mathbf{f}(\tau_{\mathrm{next}})$ of the sampling point τ_{next} and then updates the Pareto set and the training set with $(\tau_{\mathrm{next}}, \mathbf{f}(\tau_{\mathrm{next}}))$. These steps are repeated until the saturated Pareto front is obtained.

The stopping criterion of the proposed MBO-ILUTM optimizer is that the Pareto front reaches saturation state. The criterion of this state is that the new sampling points in the objective function space can not bring or only bring a small hypervolume improvement. In

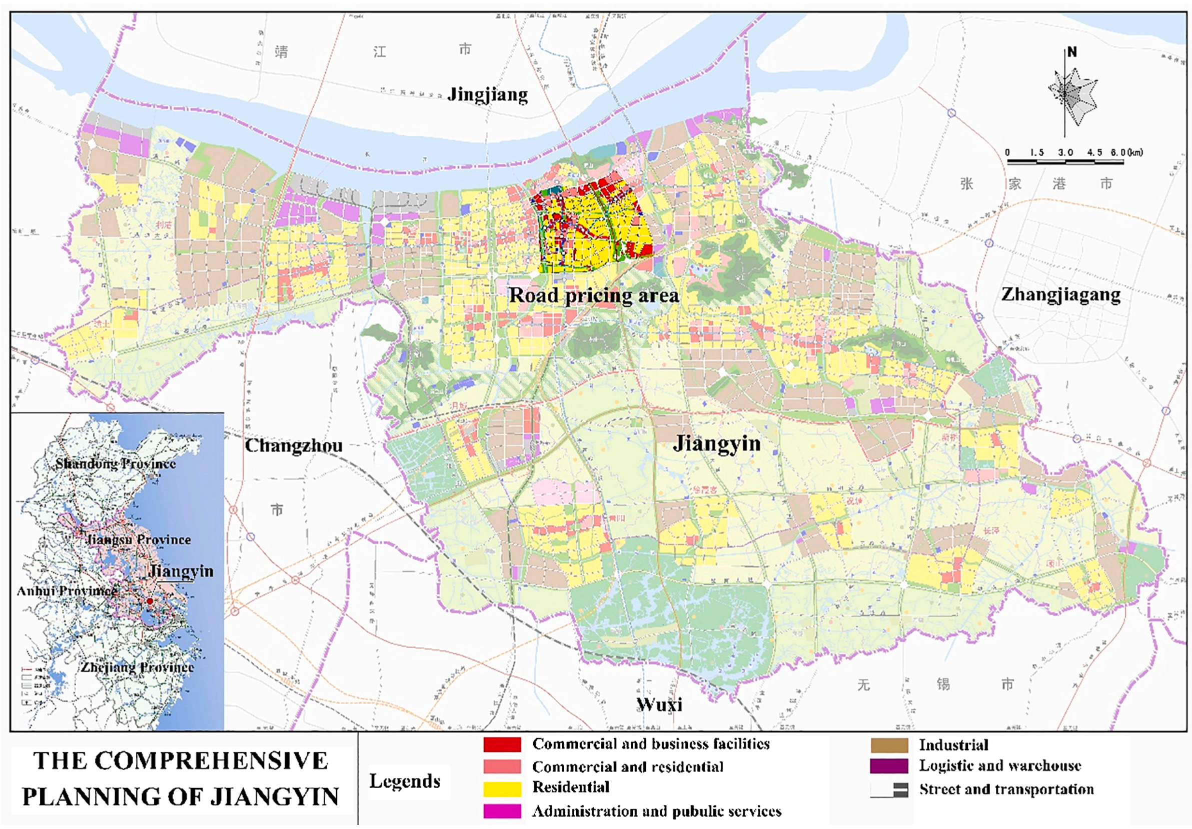

Fig. 5. Road pricing area in Jiangyin City.

the following section, the saturation state of the Pareto front will be discussed in detail.

4. Case study

4.1. Study area selection and data sources

This study selects Jiangyin City as the study area. Jiangyin City is located in the southeast of Jiangsu Province, China, with an area of about 988 square kilometers. At the end of 2018, the total population was approximately 1.65 million. As a fast-growing city, the city center of Jiangyin faces serious traffic congestion problems. This study tries to alleviate traffic congestion in Jiangyin City through the cordon-based road congestion pricing policy. This type of road congestion pricing policy is widely applied in many cities around the world, such as London, Milan, and Stockholm (Lehe, 2019). Fig. 5 presents the range of the road pricing area. Cars driving into the charging area during peak hours will be charged a certain amount of road congestion fees. Considering the per capita income level in Jiangyin City, the road congestion toll is set between 10 and 30 Chinese Yuan (CNY).

The TRANUS model is used as the lower-level integrated land use and transportation model to analyze the impacts of road congestion pricing on regional future land use and transportation systems. The basic analysis unit of the Jiangyin integrated land use and transportation model is the traffic analysis zone. Considering the size of the study area, Jiangyin City is divided into 265 TAZs, among which 56 TAZs are within the toll ring and 209 TAZs are outside of the toll ring. The initial year of the integrated land use and transportation model of Jiangyin City is 2010. The model runs every 5 years until 2030. The road congestion pricing policy is implemented in 2020. The result of the year 2030 is used to analyze the influence of road congestion pricing since the effects of road congestion pricing on regional land use need a long time to be manifested (Tillema et al., 2010; Whitehead, 2005; Zhong et al., 2015).

The data required for the integrated land use and transportation model of Jiangyin City includes two types: land use data and transportation data. The land use data is obtained from the Jiangyin comprehensive plan (2011–2030) (Jiangsu Institute of Urban Planning and Design, 2011) and Jiangyin Statistical Yearbook. The transportation data are obtained mainly from the Jiangyin Municipal Transportation Bureau and Jiangyin Public Transport Company. Besides, based on the on-the-spot investigation, the traffic flow data on the selected main roads was collected during peak hours. Based on the above data, Zhong et al., (2015) calibrated the Jiangyin integrated model using a piecewise estimation method.

4.2. Results and analysis

This subsection comprises of two parts. The effectiveness of the active learning algorithm is demonstrated in the first part, and the proposed multi-objective bi-level programming road congestion pricing model and the algorithm are applied to Jiangyin City in the second part.

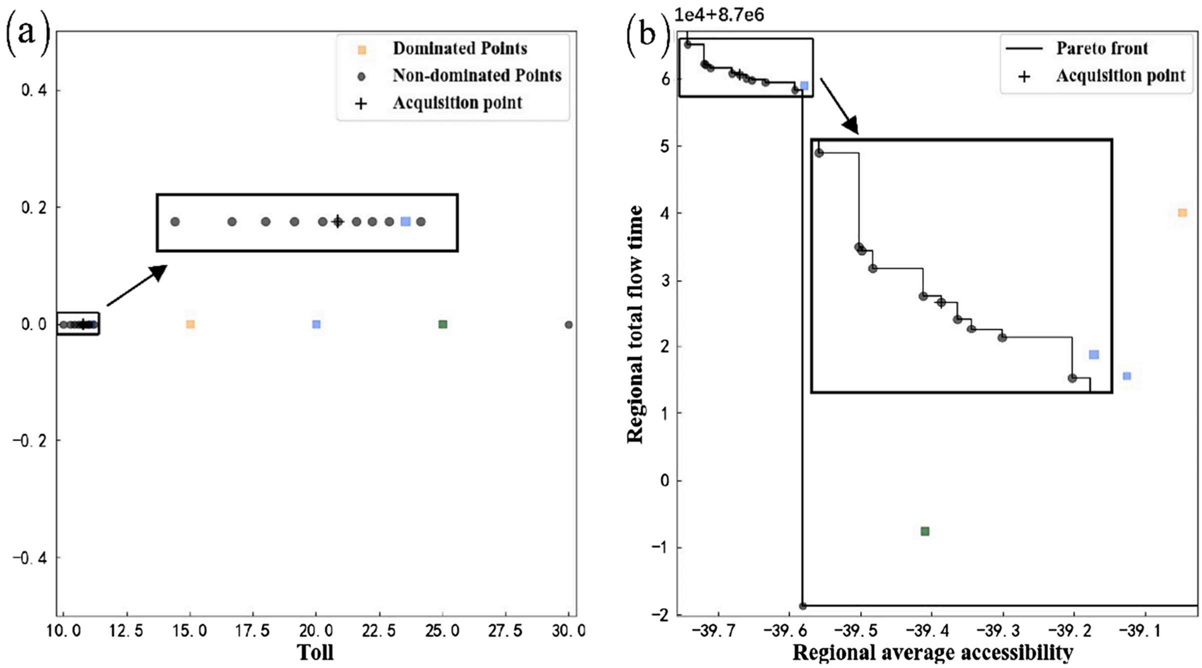

Fig. 6. (a) Solution space (b) Objective function space. Note: For the convenience of the display, the coordinate axis scale was processed utilizing numerical value plus compensation. For example, the corresponding value of the y-axis at 1 in Fig. 6 (b) is 1 * 1e4 + 8.7e6, namely, 1 * 10^4 + 8.7 * 10^6.

4.2.1. Effectiveness of the active learning algorithm

The effectiveness of the proposed active learning algorithm is assessed from the aspects of convergence speed, convergence state, and solution quality. Two objectives, the total flow time and the regional average accessibility, are considered in this experiment since it is convenient for readers to understand the changes of the Pareto front with the number of iterations more intuitively and comprehensively.

Olofsson et al. (2019) classified the Pareto front into four forms: convex, concave, both convex and concave, and discontinuous. They compared the Pareto front obtained by different algorithms with the real Pareto front, and confirmed that the multi-objective Bayesian optimization method used in this paper is superior to the existing heuristic algorithms, such as genetic algorithm, in both the speed of the algorithm and the quality of the solution.

(1) Convergence speed

At first, 5 levels of toll rates between 10 CNY and 30 CNY are randomly selected as the initial sampling points. The problem is then solved using the proposed MBO-ILUTM optimizer. The Pareto front reaches saturation state after 10 iterations, which illustrates that the proposed active learning algorithm has a fast convergence speed.

Fig. 6(a) and Fig. 6(b) show the solution space and objective function space, respectively. The circular points represent the Pareto optimal solution. The square points represent the dominated points. The higher the gray level of the dominated point, the more points dominate this point. Specifically, the dominated points have three gray levels: green, blue, and yellow, which means the point is dominated by one point, two points, and three or more points, respectively. The cruciform points represent the sampling points. In the solution space, the sampling points correspond to the points selected by the acquisition function during the current iteration. In the objective function space, the sampling points correspond to the objective function values of points selected by the acquisition function which is obtained from the integrated land use and transportation model.

Fig. 7(a) is a partially enlarged view of the objective function space in Fig. 6. Fig. 7(b) shows the changes of the two objective function values with different iteration times. The black circular points in Fig. 7(a) represent the Pareto optimal solutions. The numbers in parentheses indicate the iteration times corresponding to the point. The lines in different colors are the Pareto front obtained after the n^{th} iteration. The detailed sampling process of the MBO-ILUTM optimizer proposed in this paper can be seen in Fig. 7(a). If a new color curve appears in the lower-left corner of an existing curve, it means that the Pareto front has been updated. As the iteration times increase, the Pareto front becomes more and more detailed until it reaches saturation. It should be noted that there does not exist the 8^{th} sampling point in Fig. 7(a) because it is dominated by other points.

(2) Convergence state

The convergence state of the proposed active learning algorithm is evaluated by the hypervolume improvement calculated by Formulas (19) and (20). Fig. 8 shows the changes of the hypervolume improvement with the iteration times. It can be seen that the rate

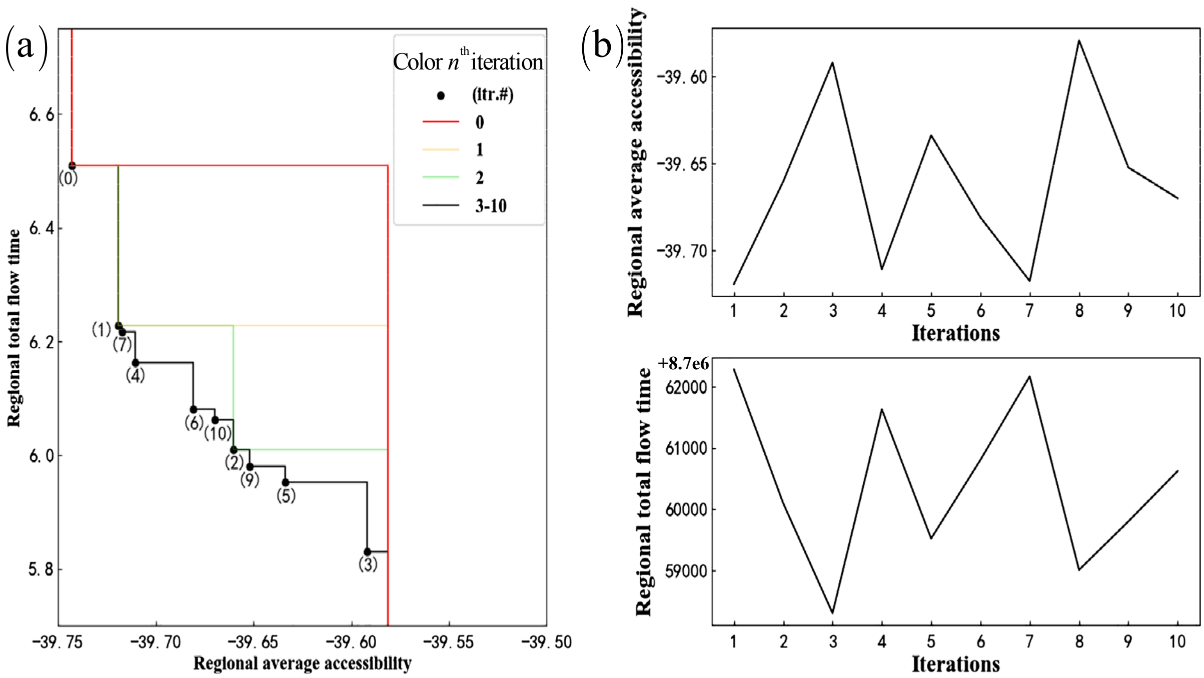

Fig. 7. (a) Partially enlarged view of the (b) Changes of objective function values with objective function space iteration times.

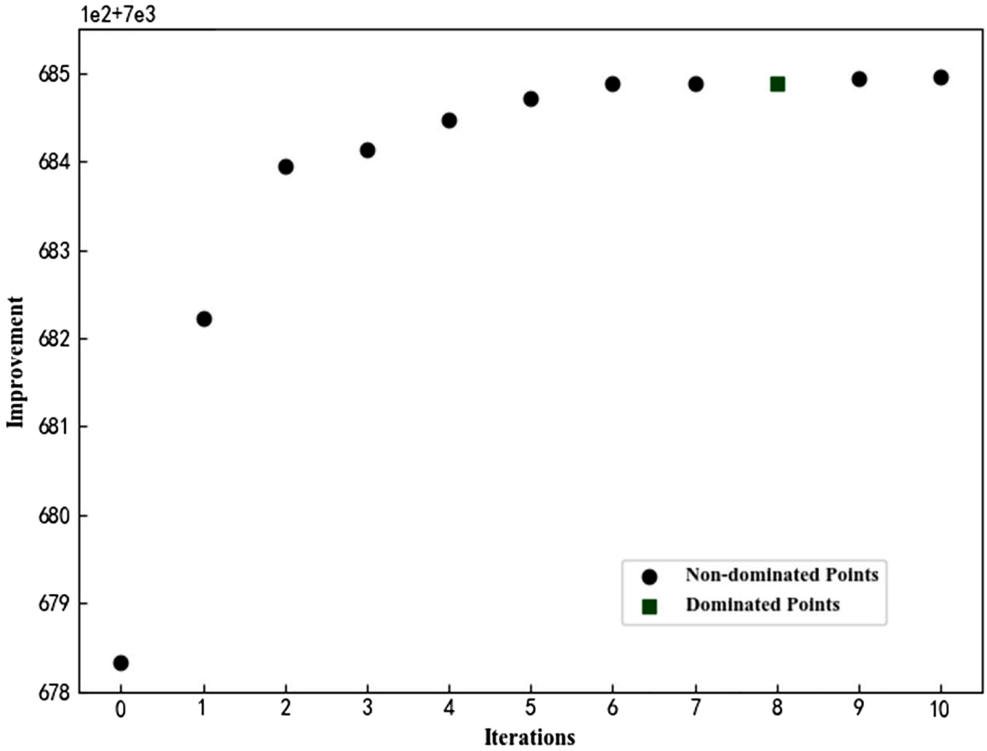

Fig. 8. Changes of the hypervolume improvement with the iteration times under two objective functions.

of increase of hypervolume improvement is very low since the 7th iteration. The hypervolume improvement becomes stable and the Pareto front reaches saturation after 10 iterations, the iteration ends. In addition, it can also be seen that 9 sampling points in the 10 iterations are Pareto optimal solutions, which is mainly benefited from the excellent data fitting ability of the Gaussian process and the maximum hypervolume improvement search strategy in Bayesian optimization.

(3) Solution quality

As mentioned earlier, the hypervolume improvement is an evaluation index for judging the convergence of the algorithm and evaluating the quality of the solution, because it represents the improvement degree of the new Pareto front compared with the existing Pareto front. For the two objective optimization problem in this section, the algorithm stops after 10 iterations. The obtained saturated Pareto front has both convex and concave sections, and the quality of the solution is good. The above conclusions are based on the following analysis and judgment:

Firstly, from the perspective of the algorithm principle, the multi-objective Bayesian optimization method provides a good balance between exploration (enhance overall accuracy) and exploitation (improve current local accuracy) (Couckuyt et al., 2014). This can ensure that the algorithm can not only jump out of the local optimal solution, but also ensure the quality of the local solution. In this example, as shown in Fig. 6 (b), the algorithm explores the upper-left corner and the lower-left corner respectively to ensure that the global optimal solution can be obtained. At the same time, after the preliminary exploration, the algorithm then conducts detailed exploitation near the upper-left corner to enhance the solution quality in this local region (see Fig. 7 (a)).

Secondly, the Pareto front obtained in the example has both convex and concave sections. The main reason is that the two objective functions, that is regional average accessibility and regional total flow time, are not always in a trade-off relationship. If they were strictly antagonistic, the Pareto front should be a continuous curve.

Finally, it can be seen from Fig. 8 that since the 7th iteration, the improvement degree obtained by updating the Pareto front tends to be stable. Since then the new sampling points can not greatly improve the quality of the solution, indicating that the solution on the Pareto front is close to an optimum, and the quality of the current solution is good. Of course, the user can choose to continue iterating to get more precise solutions, but the corresponding improvement will be small.

4.2.2. Case study

An empirical analysis is conducted based on the data of Jiangyin City. Three objectives, the regional average accessibility, the regional average land use diversity, and the regional total flow time, are taken into account. In this example, 5 sampling points are generated as the initial solutions, and the Pareto front becomes saturated after 20 iterations. Fig. 9(a) and (b) illustrate the solution space and the objective function space after 20 iterations. Fig. 9(c), (d), and (e) are the two-dimensional section view of Fig. 9(b). It

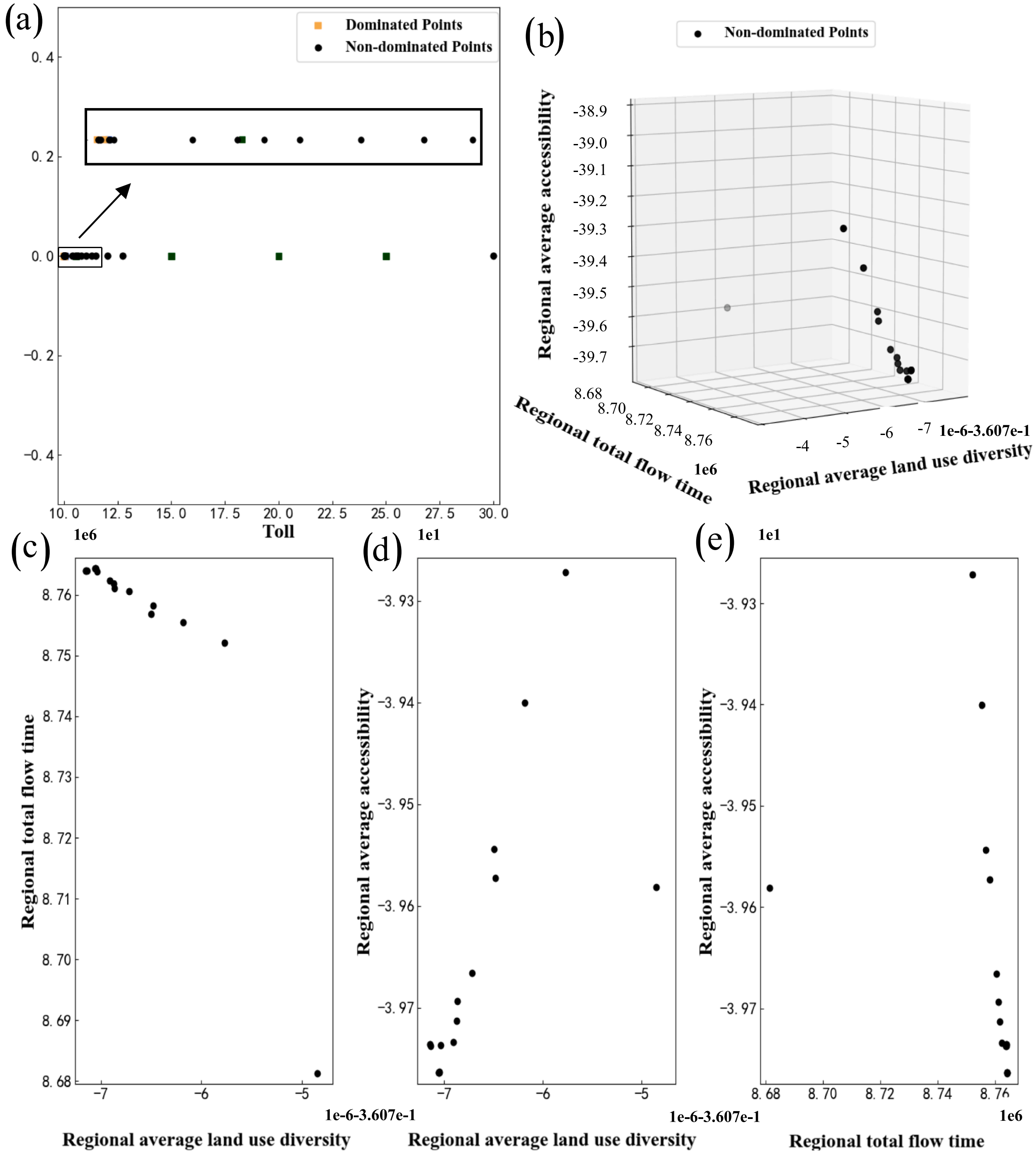

Fig. 9. (a) Solution space, (b) Objective function space, (c) Two-dimensional section view of regional total flow time and regional land use diversity, (d) Two-dimensional section view of regional average accessibility and regional land use diversity, (e) Two-dimensional section view of regional average accessibility and regional total flow time.

should be noted that due to the existence of the third objective function, the points in Fig. 9(c)–(e) do not represent the Pareto front of the two objectives.

The changes of the hypervolume improvement associated with the iteration times are presented in Fig. 10. It can be found that the hypervolume improvement increased insignificantly by updating the Pareto front since the 17th iteration and stabilizes after 20 iterations. Besides, comparing Fig. 8 and Fig. 10, it can be seen that the number of iterations required to reach the saturated Pareto front increase as the number of objective functions grows from 2 to 3.

Table 2 compares the performance of different objective functions under different road congestion pricing schemes (i.e., the traditional time-based road congestion pricing, the random road congestion pricing scheme, and the Pareto optimal road congestion

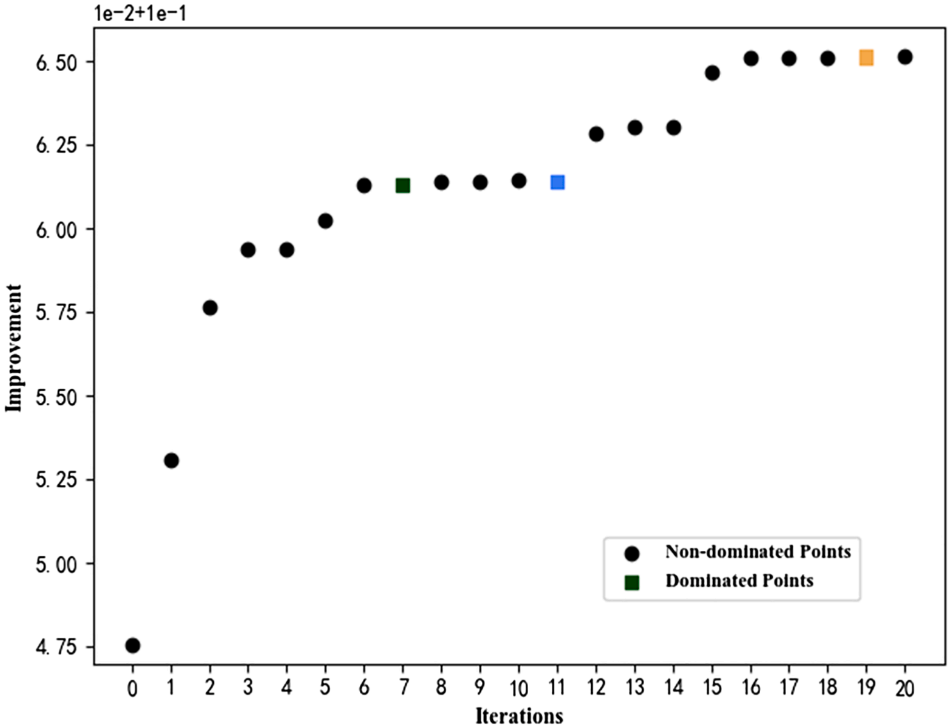

Fig. 10. Changes of the hypervolume improvement with the iteration times under three objective functions.

Table 2
Performance of different objective functions under different pricing schemes.

Road congestion pricing scheme	Toll rate (CNY)	Iterations	Regional average accessibility	Regional average land use diversity	Regional total flow time
Time-based scheme	30	/	0	0	0
Random scheme 1	15	0	−1.35%	−0.38E−06	0.68%
Random scheme 2	25	0	−0.43%	−2.50E−06	0.13%
Pareto optimal scheme	11.04	9	0.39%	6.36E−06	0.95%

pricing), where the time-based road congestion pricing is regarded as the reference scheme.

The toll rate of the time-based road congestion pricing is 30 CNY. Under this scheme, the regional total flow time is the least while the land use indicators (i.e., regional average accessibility and regional land use diversity) are not good. Two toll rates, 15 CNY and 20 CNY, are randomly selected under the random road congestion pricing scheme. Under this scheme, the regional total flow time is longer and the land use indicators are also worse than those under the time-based road congestion pricing scheme, which indicates that as the number of objective functions increases, it is difficult to find the optimal road congestion pricing scheme through scenario planning or enumeration method. According to the result of the proposed approach, one of the optimal toll rates (randomly selected from the Pareto front) is 11.04 CNY. Under this road congestion pricing scheme, the regional average accessibility and the land use diversity are improved. Although the regional total flow time is increased, the rate of increase is acceptable. This means that the proposed approach can coordinate the relationship between land use and transportation and determine the optimal toll scheme.

5. Conclusion

Although the original intention of road congestion pricing policy is to reduce the demand for car use in the toll area and mitigate traffic congestion, the negative land use effects generated by this policy have gradually attracted more and more attention. Clarifying and optimizing these negative effects are the key to the successful implementation of the road congestion pricing policy. To quantify and reduce the negative land use effect of road congestion pricing, a multi-objective bi-level programming model based on the integrated land use and transportation model is developed to optimize the road congestion pricing problem. The proposed model aims to optimize the performance of regional average accessibility, regional average land use diversity, and regional total flow time. The multi-objective bi-level programming model is NP-hard. An active learning algorithm based on multi-objective Bayesian optimization, which can reduce the times of calling the lower-level integrated land use and transportation model and improve the computation efficiency, is innovatively developed to solve the problem. A case study in Jiangyin City verifies the effectiveness of the proposed model and algorithm and illustrates that the proposed approach can be used to coordinate the relationship between land use and transportation and determine the Pareto optimal road congestion pricing scheme.

The multi-objective bi-level programming road congestion pricing model and the active learning optimization algorithm based on multi-objective Bayesian optimization proposed in this paper provide new directions for future research. First, to simplify the problem, the upper-level problem in this paper has only one decision variable (one single toll level), but the method proposed in this paper is also applicable to the problems of multiple decision variables and multiple objectives. It will be meaningful research to extend the one single toll level problem into a vector of tolls problem, such as area-based road pricing or multiple cordon charges. Second, the proposed model can be used to identify the effects of other land use and transportation policies on urban land use and transportation systems, since the lower-level problem is an integrated land use and transportation model. Third, the proposed active learning optimization algorithm can be applied to solve other black box problems with large computations. For example, by replacing the lower-level model with a traffic simulation model or an agent-based model, the proposed approach in this paper can be used to solve a series of traffic simulation and control problems, such as traffic signal control problem.

CRediT authorship contribution statement

Shaopeng Zhong: Conceptualization, Methodology, Software, Investigation, Visualization, Writing - original draft, Funding acquisition. **Yunhai Gong:** Software, Investigation, Formal analysis, Validation, Writing - original draft. **Zhijian Zhou:** Software, Investigation, Formal analysis, Validation, Writing - original draft. **Rong Cheng:** Writing - original draft. **Feng Xiao:** Funding acquisition, Supervision, Resources, Validation, Writing - review & editing.

Acknowledgement

This research has been supported by the National Natural Science Foundation of China (Project No. 71701030, 71971038), the National Science Fund for Distinguished Young Scholars (Project No.72025104), the Humanities and Social Sciences Youth Foundation of the Ministry of Education of China (Project No. 17YJCZH265), and the Fundamental Research Funds for the Central Universities of China (Project No. DUT20GJ210).

References

Allende, G.B., Still, G., 2013. Solving bilevel programs with the KKT-approach. Math. Program. 138 (1–2), 309–332.
Anas, A., 2014. The cost of congestion and the benefits of congestion pricing: a general equilibrium analysis. Transport. Res. Part B: Methodol. 136, 110–137.
Arentze, T., Timmermans, H., 2007. Congestion pricing scenarios and change of job or residential location: results of a stated adaptation experiment. J. Transp. Geogr. 15 (1), 56–61.
Bandeira, J.M., Coelho, M.C., Sá, M.E., Tavares, R., Borrego, C., 2011. Impact of land use on urban mobility patterns, emissions and air quality in a Portuguese medium-sized city. Sci. Total Environ. 409 (6), 1154–1163.
Banister, D., 2002. The integration of road pricing with land use planning. In: Proceedings of Essay for the Second Seminar of the IMPRINT-EUROPE Thematic Network: Implementing Reform on Transport Pricing: Identifying Mode-specific Issues, Brussels.
Bhat, C.R., Castelar, S., 2002. A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. Transport. Res. Part B: Methodol. 36 (7), 593–616.
Brownstone, D., Small, K.A., 2005. Valuing time and reliability: assessing the evidence from road pricing demonstrations. Transport. Res. Part A: Policy Practice 39 (4), 279–293.
Chen, P., Merrick, B.M., Brazil, T.J., 2015. Bayesian optimization for broadband high-efficiency power amplifier designs. IEEE Trans. Microw. Theory Tech. 63 (12), 4263–4272.
Chen, X., He, X., Xiong, C., Zhu, Z., Zhang, L., 2019. A Bayesian stochastic kriging optimization model dealing with heteroscedastic simulation noise for freeway traffic management. Transport. Sci. 53 (2), 545–565.
Couckuyt, I., Deschrijver, D., Dhaene, T., 2014. Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization. J. Global Optim. 60 (3), 575–594.
de la Barra, T., Pérez, B., Vera, A.N., 1984. TRANUS-J: Putting large models into small computers. Environ. Plan. B: Plan. Des. 11 (1), 87–101.
Feliot, P., Bect, J., Vazquez, E., 2017. A Bayesian approach to constrained single-and multi-objective optimization. J. Global Optim. 67 (1–2), 97–133.
Golovin, D., Solnik, B., Moitra, S., Kochanski, G., Karro, J., Sculley, D., 2017. Google vizier: A service for black box optimization. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, United States, pp. 1487–1495.
Gupta, S., Kalmanje, S., Kockelman, K.M., 2006. Road pricing simulations: Traffic, land use and welfare impacts for Austin, Texas. Transport. Plan. Technol. 29 (01), 1–23.
Jiangsu Institute of Urban Planning and Design, 2011. The Comprehensive Planning of Jiangyin (2011–2030). (<http://www.jiangyin.gov.cn/doc/2012/05/10/399237.shtml>(2019.11.11)).
Keane, A.J., 2006. Statistical improvement criteria for use in multiobjective design optimization. AIAA J. 44 (4), 879–891.
Kieffer, E., Danoy, G., Bouvry, P., Nagih, A., 2017. Bayesian optimization approach of general bi-level problems. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 1614–1621.
Lehe, L., 2019. Downtown congestion pricing in practice. Transport. Res. Part C: Emerg. Technol. 100, 200–223.
Liu, Z., Wang, S., Meng, Q., 2014. Optimal joint distance and time toll for cordon-based congestion pricing. Transport. Res. Part B: Methodol. 69, 81–97.
Lobe, P., Duchâteau, H., Stratec, S., 1998. Impacts of transport price on mobility and land use in the Brussels area. In: Proceedings of the International Symposium on Technological and Environmental Topics in Transports–"Externalities in the Urban Transport: Assessing and Reducing the Impacts.
Nahvi, A., Sadoughi, M.K., Arabzadeh, A., Sassani, A., Hu, C., Ceylan, H., Kim, S., 2019. Multi-objective Bayesian optimization of super hydrophobic coatings on asphalt concrete surfaces. J. Comput. Des. Eng. 6 (4), 693–704.
Olofsson, S., Mehrian, M., Calandra, R., Geris, L., Deisenroth, M.P., Misener, R., 2019. Bayesian multiobjective optimisation with mixed analytical and black-box functions: application to tissue engineering. IEEE Trans. Biomed. Eng. 66 (3), 727–739.
Osorio, C., Bierlaire, M., 2013. A simulation-based optimization framework for urban transportation problems. Oper. Res. 61 (6), 1333–1345.
Osorio, C., Chong, L., 2015. A computationally efficient simulation-based optimization algorithm for large-scale urban transportation problems. Transport. Sci. 49 (3), 623–636.
Park, S., Na, J., Kim, M., Lee, J.M., 2018. Multi-objective Bayesian optimization of chemical reactor design using computational fluid dynamics. Comput. Chem. Eng. 119, 25–37.

Pramanik, S., Roy, T.K., 2007. Fuzzy goal programming approach to multilevel programming problems. Eur. J. Oper. Res. 176 (2), 1151–1166.
Quddus, M.A., Carmel, A., Bell, M.G., 2007. The impact of the congestion charge on retail: The London experience. J. Transp. Econ. Policy (JTEP) 41 (1), 113–133.
Rasmussen, C.E., 2003. Gaussian processes in machine learning. In: Summer School on Machine Learning. Springer, Berlin, Heidelberg, pp. 63–71.
Rasmussen, C.E., Williams, C.K., 2006. Gaussian processes for machine learning. MIT Press, Cambridge, MA.
Rodrigues, F., Pereira, F.C., 2018. Heteroscedastic Gaussian processes for uncertainty modeling in large-scale crowdsourced traffic data. Transport. Res. Part C: Emerg. Technol. 95, 636–651.
Safirova, E., Houde, S., Lipman, D., Harrington, W. Baglino, A., 2006. Congestion Pricing: Long-term Economic and Land Use Effects, Washington, DC: Resources for the Future. Technical Report (Discussion paper 06–37).
Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N., 2015. Taking the human out of the loop: a review of Bayesian optimization. Proc. IEEE 104 (1), 148–175.
Still, B.G., May, A.D., Bristow, A.L., 1999. The assessment of transport impacts on land use: practical uses in strategic planning. Transp. Policy 6 (2), 83–98.
Sumalee, A., Shepherd, S., May, A., 2009. Road user charging design: dealing with multi-objectives and constraints. Transportation 36 (2), 167.
Sun, D., Zhang, K., Shen, S., 2018. Analyzing spatiotemporal traffic line source emissions based on massive didi online car-hailing service data. Transport. Res. Part D: Transp. Environ. 62, 699–714.
Sun Jian, D., Zhang, Y., Xue, R., Zhang, Y., 2017. Modeling carbon emissions from urban traffic system using mobile monitoring. Sci. Total Environ. 599, 944–951.
Sun, H., Gao, Z., Wu, J., 2008. A bi-level programming model and solution algorithm for the location of logistics distribution centers. Appl. Math. Model. 32 (4), 610–616.
Svenson, J., Santner, T.J., 2016. Multiobjective optimization of expensive-to-evaluate deterministic computer simulator models. Comput. Stat. Data Anal. 94, 250–264.
Szeto, W.Y., Jiang, Y., 2014. Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. Transport. Res. Part B: Methodol. 67, 235–263.
Talbi, E.G., 2013. A taxonomy of metaheuristics for bi-level optimization. In: Metaheuristics for Bi-level Optimization. Springer, Berlin, Heidelberg, pp. 1–39.
Thoai, N.V., Yamamoto, Y., Yoshise, A., 2005. Global optimization method for solving mathematical programs with linear complementarity constraints. J. Optim. Theory Appl. 124 (2), 467–490.
Tillema, T., van Wee, B., Ettema, D., 2010. The influence of (toll-related) travel costs in residential location decisions of households: a stated choice approach. Transport. Res. Part A: Policy Practice 44 (10), 785–796.
Transport for London, 2008. Central London Congestion Charging Impacts Monitoring Sixth Annual Report. London Transport Museum, London.
Wang, G., Jianq, B., Zhu, K., Wan, Z., 2010. Global convergent algorithm for the bilevel linear fractional-linear programming based on modified convex simplex method. J. Syst. Eng. Electron. 21 (2), 239–243.
Whitehead, T., 2002. Road user charging and business performance: Identifying the processes of economic change. Transp. Policy 9 (3), 221–240.
Whitehead, T., 2005. Transport charging interventions and economic activity. Transp. Policy 12 (5), 451–463.
Xiao, F., Long, J., Li, L., Kou, G., Nie, Y., 2019. Promoting social equity with cyclic tradable credits. Transport. Res. Part B: Methodol. 121, 56–73.
Xiao, F., Zhang, H.M., 2014. Pareto-improving and self-sustainable pricing for the morning commute with nonidentical commuters. Transport. Sci. 48 (2), 159–169.
Yang, H., Bell, M.G., 1997. Traffic restraint, road pricing and network equilibrium. Transport. Res. Part B: Methodol. 31 (4), 303–314.
Yang, H., Zhang, X., 2002. Multiclass network toll design problem with social and spatial equity constraints. J. Transp. Eng. 128 (5), 420–428.
Yin, Y., Lawphongpanich, S., 2009. Alternative marginal-cost pricing for road networks. Netnomics 10 (1), 77–83.
Zhang, D., Xiao, F., Shen, M., Zhong, S., 2021. DNEAT: A novel dynamic node-edge attention network for origin-destination demand prediction. Transport. Res. Part C: Emerg. Technol. 122, 102851.
Zhang, X., Yang, H., Huang, H.J., 2008. Multiclass multicriteria mixed equilibrium on networks and uniform link tolls for system optimum. Eur. J. Oper. Res. 189 (1), 146–158.
Zhong, S., Bushell, M., 2017. Built environment and potential job accessibility effects of road pricing: a spatial econometric perspective. J. Transp. Geogr. 60, 98–109.
Zhong, S., Cheng, R., Jiang, Y., Wang, Z., Larsen, A., Nielsen, O.A., 2020a. Risk-averse optimization of disaster relief facility location and vehicle routing under stochastic demand. Transport. Res. Part E: Logist. Transport. Rev. 141, 102015.
Zhong, S., Cheng, R., Li, X., Wang, Z., Jiang, Y., 2020b. Identifying the combined effect of shared autonomous vehicles and congestion pricing on regional job accessibility. J. Transp. Land Use 13 (1), 273–297.
Zhong, S., Wang, S., Jiang, Y., Yu, B., Zhang, W., 2015. Distinguishing the land use effects of road pricing based on the urban form attributes. Transport. Res. Part A: Policy Practice 74, 44–58.
Zhong, S., Wang, Z., Wang, Q., Liu, A., Cui, J., 2021. Exploring the spatially heterogeneous effects of urban built environment on road travel time variability. J. Transport. Eng., Part A: Syst. 147 (1), 4020142.